PHYS

VOLUME 94

A journal of exp

Published by the American Institute of
for the

AMERICAN PHYSICAL SOC

Officers

President

H. A. Bethe
Cornell University
Ithaca, New York

Vice-President

R. T. Birge
University of California,
Berkeley, California

Secretary

K. K. Darrow
Bell Telephone Laboratories
(after
Columbia University
New York 27, New York

Treasurer

G. B. Pegram
Columbia University
New York 27, New York

Local Secretary
for the Pacific Coast

J. Ramsay
University of Cal
Angeles
Los Angeles, California

Managing Editor

S. A. Goudsmit
Brookhaven National Labora
Upton, Long Island
New York

The Physical Review
Entered as second-class matter
Acceptance for mailing at special rate of

AL REVIEW

Established by E. L. Nichols in 1893

BOARD OF EDITORS

Associate Editors

1954–1955

W. A. Nierenberg
Herbert Goldstein
Carson Mark Kivel
R. E. Marshak
W. G. H. Panofsky
R. F. Wehrmann

Associate Editors

S. Pasternack

should be submitted to S. A. Goud
Laboratory, Upton, Long Island, New York
requested to pay a publication charge of $
author there to 100 free reprints. Extra

concerning papers in the process of publica
the Publication Manager, American Physical
, New York 22, New York.

Subscription Price

Foreign: .....
.....

Back Numbers

Price per complete year to subscribers:
1893–1913 $25.00
1914–1928 $50.00
1929 prior to July, 1939:
July, 1939–December, 1941:
January, 1942–December, 1946:
thereafter:
Price per issue $5.00

Remittance and orders for back numbers should be addressed to
Physical Review, 57 East 55 Street, New York 22, New York

This issue of members of the American Physical Society
Discount: In the case of other subscriptions

Printed by Lancaster Press, Inc., Lancaster, Pennsylvania
Lancaster, Pennsylvania, under the Act of March 3,
Section 34.64, P. L. & R. of 1948, authorized June

THE
PHYSICAL REVIEW

A journal of experimental and theoretical physics established by E. L. Nichols in 1893

SECOND SERIES, VOL. 94, NO. 2 APRIL 15, 1954

Quantum Theory of a Damped Electrical Oscillator and Noise. II. The Radiation Resistance*

J. WEBER
Glenn L. Martin College of Engineering and Aeronautical Sciences, University of Maryland, College Park, Maryland
(Received October 8, 1953)

The results of a previous paper are extended to include damping and noise due to a radiation resistance. The average electromagnetic field energy of an oscillator of natural frequency ω, with inductance L, coupled to a radiation resistance R, as a function of time t is given by

$$U=\left[\tfrac{1}{2}\hbar\omega+\frac{\hbar\omega}{\exp(\hbar\omega/kT)-1}\right][1-e^{-(R/L)t}]+U_0e^{-(R/L)t}.$$

Using first-order perturbation theory and quantum statistics, an expression is derived for the radiation resistance of a circuit. This agrees exactly with the classical value.

For an oscillator damped by a radiation resistance the vacuum fluctuation noise, and available noise power are shown to be the same as for an oscillator damped by a lumped resistance. All of the previous results are shown to be applicable to cavity resonators.

INTRODUCTION

IN a previous paper,[1] hereafter denoted by I, we discussed some quantum effects in damping and noise. The vacuum fluctuations were shown to be directly observable in noise experiments. The assumption was made that the circuit did not radiate. In a subsequent paper it will be shown that the vacuum fluctuations can be observed as noise induced in an electron stream which interacts with a circuit. To do such an experiment coupling holes would have to be provided, and a radiation resistance would be thereby introduced. For this reason the results of I are extended to include the radiation resistance. It will be shown that the radiation resistance affects the system in essentially the same way as an ordinary resistance. Also for some purposes a cavity resonator is more suitable than a circuit composed of lumped elements. It will therefore be shown that the results already obtained are applicable to cavity resonators also.

* Supported by the U. S. Office of Naval Research.

[1] J. Weber, Phys. Rev. **90**, 977 (1953). We have used the notation of the previous paper except that in this paper the variable q of the circuit is written as q_F.

ELECTRICAL OSCILLATOR DAMPED BY RADIATION RESISTANCE

We assume that the radiation resistance is a series element, as in Fig. 1. For the Hamiltonian of Fig. 1, we have

$$\mathcal{H}=\tfrac{1}{2}(p_F^2+\omega^2q_F^2)+H_{FR}+H_s'+\textstyle\sum_\lambda 1/2m_\lambda(\mathbf{p}_\lambda-e_\lambda\mathbf{A}/c)^2. \quad (1)$$

p_F and q_F are the field variables employed in I, and the first two terms of (1) are the Hamiltonian of the dissipationless oscillator. The term H_{FR} is the Hamiltonian of the unperturbed radiation fields, and the term H_s' represents part of the unperturbed Hamil-

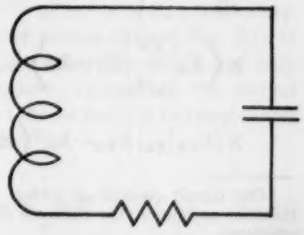

FIG. 1. Electrical oscillator damped by a radiation resistance.

tonian of the electrons.† The last term is a summation over all of the electrons, and represents the interaction between the electrons and the circuit and radiation fields. m_λ is the mass, $\mathbf{p}_\lambda$ is the momentum, and e_λ is the charge for the λth electron, and $\mathbf{A}$ is the magnetic vector potential of the radiation fields. We assume that no resistance other than radiation resistance is present, for simplicity. If both kinds of resistance are present additional terms are added to the Hamiltonian and the effects can be shown to take place independently.

We consider our oscillator to be weakly damped, and employ first-order perturbation theory. The term in A^2 in (1) can be neglected. The Hamiltonian becomes

$$\mathcal{H}=\tfrac{1}{2}(p_F^2+\omega^2 q_F^2)+H_{FR}+H_e-\sum_\lambda \frac{e_\lambda \mathbf{p}_\lambda\cdot\mathbf{A}}{m_\lambda c}. \quad (2)$$

We have included the term $\sum_\lambda(p_\lambda^2/2m_\lambda)$ in the term H_e. We expand the radiation fields in normal modes, $\mathbf{A}=\sum_{\omega_i}\mathbf{A}_{\omega_i}(\mathbf{r})q_{\omega_i}(t)$. The last term in (2) can then be written

$$\sum \frac{e_\lambda \mathbf{p}_\lambda\cdot\mathbf{A}}{m_\lambda c}=\sum_{\omega_i}\sum_\lambda \frac{e_\lambda \mathbf{p}_\lambda\cdot\mathbf{A}_{\omega_i}q_{\omega_i}}{m_\lambda c}. \quad (3)$$

For a circuit this becomes

$$\sum_{\omega_i}\sum_\lambda \frac{e_\lambda \mathbf{p}_\lambda\cdot\mathbf{A}_{\omega_i}q_{\omega_i}}{m_\lambda c}=\int_c \frac{I(\mathbf{r})}{c}\sum_{\omega_i}\mathbf{A}_{\omega_i}q_{\omega_i}\cdot d\mathbf{l}. \quad (4)$$

We can write the current $I(\mathbf{r})$ as $I(\mathbf{r})=I_0 f(\mathbf{r})$, where $f(\mathbf{r})$ is the current distribution function and I_0 is the current at some reference point. We define F_{ω_i} by

$$F_{\omega_i}=\left|\int_c f(\mathbf{r})\mathbf{A}_{\omega_i}\cdot d\mathbf{l}\right|,$$

and we define the inductance L by the relation $LI_0^2=q_F^2\omega^2$. By utilizing these relations, Eq. (4) can be written:

$$\sum_{\omega_i}\sum_\lambda \frac{e_\lambda \mathbf{p}_\lambda\cdot\mathbf{A}_{\omega_i}q_{\omega_i}}{m_\lambda c}=\frac{\omega q_F}{c\sqrt{L}}\sum_{\omega_i}q_{\omega_i}F_{\omega_i}. \quad (5)$$

The term (5) is an interaction term which will cause transitions with exchange of energy between the circuit and the radiation fields through the coupling furnished by the electrons. The transition probability can be shown to be

$$W_r=\frac{2\pi\omega^2}{c^2\hbar}(F_\omega^2)_{Av}\Big[\rho(E_{FR}+\hbar\omega)\langle E_{FR}|q_\omega|E_{FR}+\hbar\omega\rangle^2 \times\left\langle E_F\left|\frac{q_F}{\sqrt{L}}\right|E_F-\hbar\omega\right\rangle^2+\rho(E_{FR}-\hbar\omega) \times\langle E_{FR}|q_\omega|E_{FR}-\hbar\omega\rangle^2\left\langle E_F\left|\frac{q_F}{\sqrt{L}}\right|E_F+\hbar\omega\right\rangle^2\Big]. \quad (6)$$

† Our circuit consists of perfect conductors in series with a radiation resistance. The electrons referred to are in the radiation resistance.

In Eq. (6), $(F_\omega^2)_{Av}$ is an average over all directions of $(F_\omega)^2$, the symbol $\langle E_{FR}|q_\omega|E_{FR}+\hbar\omega\rangle$ indicates the matrix element of the operator q between the quantum states of the radiation fields with eigenvalues E_{FR} and $E_{FR}+\hbar\omega$. $\langle E_F|q_F|E_F+\hbar\omega\rangle$ has the corresponding meaning for the fields of the circuit. $\rho(E_{FR}+\hbar\omega)$ and $\rho(E_{FR}-\hbar\omega)$ are the density in energy of the quantum states of the radiation fields in the vicinity of $(E_{FR}+\hbar\omega)$ and $(E_{FR}-\hbar\omega)$, respectively. We assume that initially the circuit is in an eigenstate, but that only the temperature of the radiation fields is known. We therefore need to average Eq. (6) over an ensemble of similar systems. Now the density in energy of the quantum states of the radiation fields depends only on the frequency and not on the energy, so that $\rho(E_{FR}+\hbar\omega)=\rho(E_{FR}-\hbar\omega)=\omega^2/2\pi^2\hbar c^3$. By employing the harmonic oscillator matrix elements and the relation $E_{FR}=(m+\frac{1}{2})\hbar\omega$, the average of the squared matrix element $\langle E_{FR}|q_\omega|E_{FR}+\hbar\omega\rangle^2$ can be shown to be

$$\frac{\sum_i\langle E_{FRi}|q_\omega|E_{FRi}+\hbar\omega\rangle^2\exp(-E_{FRi}/kT)}{\sum_i\exp(-E_{FRi}/kT)} =\frac{\sum_0^\infty \hbar(m+1)\exp[-(m+\frac{1}{2})\hbar\omega/kT]}{2\omega\sum_0^\infty\exp[-(m+\frac{1}{2})\hbar\omega/kT]} =\frac{\hbar}{2\omega[1-\exp(-\hbar\omega/kT)]}. \quad (8)$$

We can calculate the ensemble average of the squared matrix element $\langle E_{FR}|q_\omega|E_{FR}-\hbar\omega\rangle^2$ in the same way and employ Eq. (8) and Eq. (7) to write Eq. (6) in the form

$$W_r=\frac{\omega^3}{2\pi\hbar c^5}(F_\omega^2)_{Av}\left[\frac{1}{1-\exp(-\hbar\omega/kT)}\right] \times\left[\left\langle E_F\left|\frac{q_F}{\sqrt{L}}\right|E_F-\hbar\omega\right\rangle^2 +\left\langle E_F\left|\frac{q_F}{\sqrt{L}}\right|E_F+\hbar\omega\right\rangle\exp(-\hbar\omega/kT)\right]. \quad (9)$$

By inserting the harmonic oscillator matrix elements for the circuit, Eq. (9) becomes

$$W_r=\frac{\omega^2}{4L\pi c^5}(F_\omega^2)_{Av}\left[\frac{1}{1-\exp(-\hbar\omega/kT)}\right] \times[n+(n+1)\exp(-\hbar\omega/kT)]. \quad (10)$$

In (10) the first term of the last factor represents the effect of a downward transition and the second

term represents the effect of an upward transition. The rate of change of the field energy associated with the circuit is the difference of the two terms multiplied by $\hbar\omega$ and is

$$\frac{dU}{dt}=\frac{\hbar\omega^3}{4L\pi c^5}(F_\omega^2)_{Av}\left[\frac{1}{1-\exp(-\hbar\omega/kT)}\right]\times[(n+1)\exp(-\hbar\omega/kT)-n]. \quad (11)$$

For an ensemble, expression (11) holds for all time; we can integrate Eq. (11) to obtain

$$U=\left[\tfrac{1}{2}\hbar\omega+\frac{\hbar\omega}{\exp(\hbar\omega/kT)-1}\right]\times\left[1-\exp\left(-\frac{\omega^2}{4L\pi c^5}[(F_\omega^2)_{Av}]t\right)\right]+U_0\exp\left(-\frac{\omega^2}{4L\pi c^5}[(F_\omega^2)_{Av}]t\right). \quad (12)$$

For large energy (classical limit) the second term is the only significant one. Comparing (10) with the well-known classical result $U=U_0e^{-(R/L)t}$, we obtain

$$R=\frac{\omega^2}{4\pi c^5}(F_\omega^2)_{Av}=\frac{\omega^2}{4\pi^2c^5}\left[\left|\int_c f(\mathbf{r})\mathbf{A}_\omega\cdot\mathbf{dl}\right|^2\right]_{Av} \quad (13)$$

(13) is a formula for the radiation resistance, it agrees exactly with the classical[2] value. In terms of Eq. (13), Eq. (12) is

$$U=\left[\tfrac{1}{2}\hbar\omega+\frac{\hbar\omega}{\exp(\hbar\omega/kT)-1}\right][1-e^{-(R/L)t}]+U_0e^{-(R/L)t}. \quad (12A)$$

EQUIVALENCE OF RADIATION RESISTANCE AND A NOISE GENERATOR

In I we showed the equivalence of a noise current generator and a conductance. In this section, we will show the equivalence of a noise voltage generator and a radiation resistance. We imagine that the radiation resistance is removed and replaced by a voltage generator. In order that the transition probability be proportional to time the generator needs to have a continuous spectrum in the vicinity of ω. The Hamiltonian of the system of Fig. 2 is

$$H=\tfrac{1}{2}(p_F^2+\omega^2q_F^2)-(p_FV(t)/\omega\sqrt{L}); \quad (14)$$

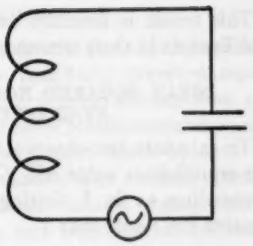

FIG. 2. Electrical oscillator and a voltage generator.

the transition probability is

$$W_G=\frac{\pi V^2(\omega)}{\omega^2\hbar^2L}[\langle E_F|p_F|E_F+\hbar\omega\rangle^2+\langle E_F|p_F|E_F-\hbar\omega\rangle^2]. \quad (15)$$

In Eq. (15) the mean-square value of the voltage over a range $d\omega$ is $V^2(\omega)d\omega$. Inserting the harmonic oscillator matrix elements, Eq. (15) becomes

$$W_G=\frac{\pi V^2(\omega)}{2\hbar\omega L}[(n+1)+n]. \quad (16)$$

In order to compare the transition probability induced by the voltage generator with that induced by the radiation resistance we employ Eq. (13) to write Eq. (10) in terms of the resistance. With this substitution, Eq. (10) becomes

$$W_r=\frac{R}{L}\left[\frac{(n+1)+n}{\exp(\hbar\omega/kT)-1}+n\right]. \quad (17)$$

Comparison of Eq. (16) and Eq. (17) shows that the transition probability will be the same insofar as the first term of Eq. (17) is concerned if

$$V^2(\omega)=\frac{2R\hbar\omega}{\pi[\exp(\hbar\omega/kT)-1]}. \quad (18)$$

Expression (18) is the Nyquist formula in the voltage representation, modified for quantum effects. There is still the last term in Eq. (17). This term is seen to be the transition probability at $T=0$, that is, the transition probability if the resistance is in its lowest state and the quantum number of the circuit is n. We conclude that for the radiation resistance the transitions required by Eq. (17) will be produced by a noise voltage generator described by Eq. (18), plus spontaneous emission, that is, plus the effect of the absorber in its lowest state.

We can also imagine the second term of Eq. (17) to be equivalent to a voltage generator which can only induce downward transitions; comparing the second term of Eq. (17) and Eq. (16) we see that the equivalent voltage for such a generator is

$$V^2(\omega)=\frac{2}{\pi}\hbar\omega R. \quad (19)$$

[2] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), p. 251. In calculating the average of (F_ω^2) two polarizations need to be included for each direction of propagation $\mathbf{A}_\omega$.

This result is formally analogous to that of[3] Park and Epstein in their treatment of spontaneous emission.

MEAN SQUARED NOISE VOLTAGE AND AVAILABLE POWER

To calculate the observable noise voltage we obtain the equilibrium value of $[\int \mathbf{E}\cdot d\mathbf{l}]^2$ over the ensemble. Proceeding as in I, letting C be the capacity, one obtains the result that

$$\langle V^2\rangle_{h\nu}=\frac{\sum_n \left\langle E_{Fn}\left|\frac{\hat{p}^2}{C}\right|E_{Fn}\right\rangle \exp[-(n+\frac{1}{2})(\hbar\omega/kT)]}{\sum_n \exp[-(n+\frac{1}{2})(\hbar\omega/kT)]}. \quad (20)$$

Equation (20) is identical with Eq. (27) of I, and the value of $\langle V^2\rangle_{h\nu}$ is

$$\langle V^2\rangle_{h\nu}=\frac{1}{C}\left[\frac{\hbar\omega}{2}+\frac{\hbar\omega}{\exp(\hbar\omega/kT)-1}\right]$$
$$=\omega^2 L\left[\frac{\hbar\omega}{2}+\frac{\hbar\omega}{\exp(\hbar\omega/kT)-1}\right]. \quad (21)$$

The first term of Eq. (21) represents the effect of the vacuum fluctuations and the second term represents the effect of the thermal fluctuations. In I we proved that the first term of Eq. (21) cannot be removed by making formal changes in the Hamiltonian which remove the zero point energy, and the same proof applies here. This is because such changes in the Hamiltonian do not affect the wave functions and therefore do not affect the summations in Eq. (20).

In order to calculate the available power, we consider an arrangement similar to Fig. 1, but with two series resistances R_1 at temperature T_1 and R_2 at temperature T_2. The Hamiltonian is

$$H=\tfrac{1}{2}(p_F^2+\omega^2 q_F^2)+H_{FK}+H_e$$
$$+\frac{\omega q_F}{\sqrt{L}}[\sum_i q_{\omega i}[F_{1\omega i}+F_{2\omega i}]]. \quad (22)$$

The rate of change of field energy can be obtained in the same manner as Eq. (11) was obtained, the result is

$$\frac{dU}{dt}=\frac{\hbar\omega^3}{4L\pi c^5}\left[(F_{1\omega}^2)_{h\nu}\left[\frac{1}{1-\exp(-\hbar\omega/kT_1)}\right]\right.$$
$$\times[(n+1)\exp(-\hbar\omega/kT_1)-n]$$
$$+(F_{2\omega}^2)_{h\nu}\left[\frac{1}{1-\exp(-\hbar\omega/kT_2)}\right]$$
$$\left.\times[(n+1)\exp(-\hbar\omega/kT_2)-n]\right]. \quad (23)$$

[3] D. Park and H. T. Epstein, Am. J. Phys. **17**, 301 (1949).

By making use of Eq. (13), Eq. (23) becomes

$$\frac{dU}{dt}=\frac{\hbar\omega}{L}\left[\left(\frac{R_1}{1-\exp(-\hbar\omega/kT_1)}\right)\right.$$
$$\times((n+1)\exp(-\hbar\omega/kT_1)-n)$$
$$+\left(\frac{R_2}{1-\exp(-\hbar\omega/kT_2)}\right)$$
$$\left.\times((n+1)\exp(-\hbar\omega/kT_2)-n)\right]. \quad (24)$$

In order to calculate the noise power transferred to the system by R_1 we first determine the stationary value of n by setting Eq. (24) equal to zero, the value is

$$n=\left(\frac{R_1}{\exp(\hbar\omega/kT_1)-1}+\frac{R_2}{\exp(\hbar\omega/kT_2)-1}\right)\Big/(R_1+R_2). \quad (25)$$

This is the most probable value of n a long time after R_1 and R_2 (at temperatures T_1 and T_2) have been coupled to the circuit. We calculate the power transferred to the system by R_1 by inserting the value of n given by Eq. (25) into expression (11) for the power transferred by R_1 when the quantum number of the circuit is n. The result is

$$P_{\text{transferred by } R_1}=\frac{R_1\hbar\omega}{L}\left[\frac{1}{\exp(\hbar\omega/kT_1)-1}\right.$$
$$\left.-\left(\frac{R_1}{\exp(\hbar\omega/kT_1)-1}+\frac{R_2}{\exp(\hbar\omega/kT_2)-1}\right)\Big/(R_1+R_2)\right]. \quad (26)$$

Equation (26) will approach a maximum if $R_2/R_1\to\infty$, $T_2\to 0$; the maximum value which is approached is

$$P_{\max}\to\frac{R_1\hbar\omega}{L}\left[\frac{1}{\exp(\hbar\omega/kT_1)-1}\right]=\frac{\hbar\omega\Delta\omega}{\exp(\hbar\omega/kT_1)-1}, \quad (27)$$

and this is the same as expression (34) of I.

CAVITY RESONATORS

The results of I and this paper, may be extended to apply to a single mode of a resonant cavity in the following way. For the dissipationless cavity the Hamiltonian for a single mode is identical with that for a circuit, and the interaction terms which bring in the dissipation will be the same in form. A study of the theory developed here shows that for any mode we can define an equivalent capacity C by the relation

$$U_{\text{electric}}=\int\frac{E^2}{8\pi}dv=\tfrac{1}{2}C\left[\int_c \mathbf{E}\cdot d\mathbf{l}\right]^2. \quad (28)$$

The value of C obtained by Eq. (28) can be used in the noise formulas to calculate $\langle V^2\rangle_{kw}=[\int_c \mathbf{E}\cdot d\mathbf{l}]^2$, over the same contour which appears in expression (28). For the quantities G/C and R/L which appear in (19a) of I and (12A) of this paper we substitute the ratio,

$$\frac{\omega}{2\pi}\times\frac{\text{Energy dissipated per cycle}}{\text{Maximum stored energy per cycle}},$$

and compute this ratio classically.

CONCLUSION

Expressions (21) and (27) are the same as expressions (28) and (34) of I. We see that for a circuit damped by a radiation resistance the vacuum fluctuations are observable in electromotive force measurements at low temperatures. The available power tends to zero as the temperature approaches zero. The radiation resistance is seen to have the same classical and quantum effects as regards damping and noise, as an ordinary resistance. To first order there are no additional terms in the radiation resistance formula due to quantum effects.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Vacuum Fluctuation Noise*†

J. WEBER
Glenn L. Martin College of Engineering and Aeronautical Sciences, University of Maryland, College Park, Maryland
(Received October 8, 1953)

An electron stream interacting with a damped oscillator is considered. The vacuum fluctuations and the thermal fluctuations can be observed in the noise. It is shown that an electron stream provides a means for precisely measuring the mean squared electromotive force for certain modes.

INTRODUCTION

IN previous papers,[1,2] to be referred to as I and II, we have shown that for a damped oscillator the results of precise measurements of the mean squared electromotive force are given by

$$\langle V^2\rangle_{kw}=\frac{1}{C}\left[\frac{\hbar\omega}{2}+\frac{\hbar\omega}{\exp(\hbar\omega/kT)-1}\right]. \quad (1)$$

In Eq. (1), C is the capacity, and ω is the natural (angular) frequency of the oscillator. The first term of Eq. (1) represents the effect of the vacuum fluctuations, and the second term represents the effect of the thermal fluctuations. We consider here the possibility of observing these small fluctuations by allowing an electron stream to interact with the oscillator and observing the resultant electron stream noise.

It is well known[3] that the use of electrons as test charges does not in general lead to precise field measurements. We will show, however, that in the experiments to be discussed a measurement of the mean squared electromotive force can give precisely the value above.

INTERACTION OF AN ELECTRON STREAM WITH A DAMPED ELECTRICAL OSCILLATOR

Consider an electron stream which interacts with a damped electrical oscillator (Fig. 1).

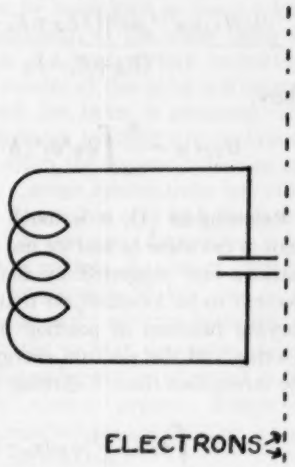

FIG. 1. An electron stream interacting with a damped oscillator.

The interaction can be imagined to take place by sending the stream of electrons near the condenser plates, through holes in the condenser plates, or through

* A brief report of this work was given at the June, 1953, Rochester Meeting of the American Physical Society.
† Supported by the U. S. Office of Naval Research.
[1] J. Weber, Phys. Rev. **90**, 977 (1953).
[2] J. Weber, preceding paper, Phys. Rev. **94**, 211 (1954).
[3] W. Heitler, *The Quantum Theory of Radiation* (Oxford University Press, London, 1944), p. 78.

a cavity resonator. The circuit is assumed to be in thermal equilibrium with the conductance which is maintained at temperature T. If the damping is small, the effect of the conductance is mainly to determine the average electromagnetic field energy of the circuit. The wave-functions for the quantized fields of the circuit will not be significantly different from those of the undamped oscillator. Under these conditions we can employ the following Hamiltonian to discuss the interaction between an electron and the fields associated with the circuit, or a single mode of a cavity resonator:

$$\mathcal{H}=P^2/2m+\tfrac{1}{2}(p^2+\omega^2q^2)-(e/mc)\mathbf{A}\cdot\mathbf{P}. \quad (2)$$

$\mathbf{A}$ is the magnetic vector potential, and we let $\mathbf{A}=\mathbf{A}_0(r)q(t)$. The variable p is canonically conjugate to q. $\mathbf{P}$ is the operator corresponding to the electron momentum. It is unnecessary to include terms in (2) representing the conductance because there is no direct interaction between the electron and the conductance.

We assume that the circuit is in an eigenstate of its unperturbed Hamiltonian before the interaction begins. For the perturbed wave functions of the system we assume the expression

$$\Psi=\textstyle\sum_{ij} a_i\psi_i b_j\phi_j \exp[-(i/\hbar)(E_i+E_j)t], \quad (3)$$

where ψ_i is an unperturbed wave function for the fields of the circuit, ϕ_j is an unperturbed wave function for the electron.

At the time interaction begins: $a_n=1$, $b_m=1$, $a_i=0$, $i\neq n$, $b_j=0$, and $j\neq m$.

It can be shown[4] that at any time t during the interaction time,

$$|a_K(t)b_l(t)|^2=\frac{4|H_{KlNM}'|^2\sin^2\{[(E_K+E_l-E_N-E_M)/2\hbar]t\}}{(E_K+E_l-E_N-E_M)^2}, \quad (4)$$

where

$$H_{KlNM}'=\frac{e}{mc}\int\psi_K^*\phi_l^*(\mathbf{A}\cdot\mathbf{P})\psi_N\phi_M d\tau_\psi d\tau_\varphi. \quad (5)$$

Referring to (4), it is much easier to calculate a_K than to calculate b_l, and we use a method and approximations first suggested by Smith.[5] We consider the electron to be localized, so that $\mathbf{A}\cdot\mathbf{P}$ is a very slowly varying function of position over the region of the electron and the electron energy is not known during the interaction time. Following Smith[5] we write

$$\int\phi_l^*(\mathbf{A}\cdot\mathbf{P})\phi_M d\tau_\varphi\approx(\mathbf{A}\cdot\mathbf{P})_M\delta_{lM},$$

and expression (5) becomes

$$H_{KlNM}'\approx H_{KN}'P_M\delta_{lM}=A_0P_M\delta_{lM}\langle E_{FK}|q|E_{FN}\rangle\frac{e}{mc}. \quad (5A)$$

In (5A), $\langle E_{FK}|q|E_{FN}\rangle$ is the matrix element of q between the quantum states of the field with quantum numbers K and N. P_M is the average value of the electron momentum over the initial and final quantum states of the electron. We assume that there are electrodes in the system beyond the interaction region which limit the electrons observed to those with velocities fairly close to the original electron velocity, and that the electron velocities are ultimately observed when the electrons have travelled a considerable distance beyond the circuit interaction gap. Making use of (5A) and the initial condition $b_m(t)=1$, we can write (4) in the form

$$|a_K(t)|^2=\frac{4e^2A_0^2P_M^2|\langle E_{FK}|q|E_{FN}\rangle|^2\sin^2(\frac{1}{2}\omega_{KN}t)}{m^2c^2\hbar^2\omega_{KN}^2}. \quad (6)$$

In (6) ω_{KN} is the natural frequency of the oscillator, and (6) gives us the probability that the field has gained or lost a quantum at time t. It is interesting to note that (6) can also be obtained directly from the Hamiltonian (2), if we quantize the field but treat the electron classically, by not regarding $\mathbf{P}$ as an operator.

We assume that the interaction gap is small and of length l. The interaction time τ will be given approximately by

$$\tau=lm/|P_M|. \quad (7)$$

From Eq. (5) of I we have

$$\langle E^2\rangle=(1/c^2)p^2\langle A_0^2\rangle, \quad p^2=CV^2=C(El)^2. \quad (8)$$

For a resonator, (8) defines C.

Making use of 6, 7, and 8, we can write Eq. (4) at time τ in the form

$$|a_K(\tau)|^2=\frac{e^2}{\hbar^2C}|\langle E_{FK}|q|E_{FN}\rangle|^2\left(\frac{\sin(\frac{1}{2}\omega_{KN}\tau)}{\frac{1}{2}\omega_{KN}\tau}\right)^2. \quad (9)$$

We let $\omega_{KN}\tau=\theta$, where θ is the electron transit angle. Making use of the matrix elements for the quantum states of the field we obtain for (9)

$$|a_K(\tau)|^2=\frac{e^2}{\hbar\omega C}\left[\frac{\sin\frac{1}{2}\theta}{\frac{1}{2}\theta}\right]^2\left(\frac{N+1}{2}\right), \quad \text{for } K=N+1, \quad (10)$$

$$|a_K(\tau)|^2=\frac{e^2}{\hbar\omega C}\left[\frac{\sin\frac{1}{2}\theta}{\frac{1}{2}\theta}\right]^2\left(\frac{N}{2}\right), \quad \text{for } K=N-1, \quad (11)$$

$$|a_K(\tau)|^2=0, \quad \text{for } K\neq N\pm1. \quad (12)$$

Expression (12) states that the exchange of energy takes place in one-quantum steps. Equation (10) is the probability that the field will gain a quantum from

[4] W. Heitler, *The Quantum Theory of Radiation* (Oxford University Press, London, 1944), p. 88.
[5] L. P. Smith, Phys. Rev. 69, 195 (1946).

the electron, and (11) is the probability that the field will lose a quantum to the electron. We assume that the experiment is so arranged that (9) is small and therefore only single-quantum exchanges are significant.

Suppose that M electrons interact with the circuit, and that a sufficiently long time elapses between interactions so that the circuit can return to a state of equilibrium with the conductance. This will be the case if the current in the electron stream is very small. Under these conditions we will have $|a_K(\tau)|^2_{K=N+1}M$ electrons lose a quantum, and $|a_K(\tau)|^2_{K=N-1}M$ electrons gain a quantum. If an electron neither gains nor loses energy we can say that the electromotive force of the circuit during that interaction time was zero. If an electron gains or loses a quantum we can say that the electromotive force during that interaction time was $V=\hbar\omega/e$. The mean squared noise voltage for a circuit whose quantum number is N is

$$\langle V^2\rangle(N)=\frac{\sum_i V_i^2}{M}$$

$$=\frac{\hbar^2\omega^2}{e^2M}[|a_K(\tau)|^2_{K=N+1}+|a_K(\tau)|^2_{K=N-1}]M$$

$$=\frac{\hbar\omega}{C}(N+\tfrac{1}{2})\left(\frac{\sin\frac{1}{2}\theta}{\frac{1}{2}\theta}\right)^2. \tag{13}$$

If we consider an ensemble, the ergodic theorem guarantees that the measured value at temperature T will be the ensemble average. The ensemble average of N is

$$N_{\hbar\omega}=\frac{\sum_N N\exp[-(N+\frac{1}{2})\hbar\omega/kT]}{\sum_N\exp[-(N+\frac{1}{2})\hbar\omega/kT]}$$

$$=\frac{1}{\exp(\hbar\omega/kT)-1}. \tag{14}$$

Inserting (14) into (13) we obtain

$$\langle V^2\rangle_{\hbar\omega}=\frac{1}{C}\left[\frac{\hbar\omega}{2}+\frac{\hbar\omega}{\exp(\hbar\omega/kT)-1}\right]\left(\frac{\sin\frac{1}{2}\theta}{\frac{1}{2}\theta}\right)^2. \tag{15}$$

If the transit angle θ is small, (15) becomes

$$\langle V^2\rangle_{\hbar\omega}=\frac{1}{C}\left[\frac{\hbar\omega}{2}+\frac{\hbar\omega}{\exp(\hbar\omega/kT)-1}\right]. \tag{16}$$

A circuit, and a cavity, will have many modes. An experiment can be so arranged that the only electrons studied are those with energies between values $E_{\max}$ and $E_{\min}$. This limits the number of modes considered to a definite number, namely those of frequency lower than $\omega_{\max}=(E_{\max}-E_{\min})/\hbar$, because we are considering one-quantum processes. The modes of frequency higher than the lowest mode can be made to have a very small contribution because the transit angle will be large and the factor $(\sin\frac{1}{2}\theta/\frac{1}{2}\theta)^2$ appearing in (15) will be very small.

Accordingly we say that the contribution of all modes other than the principal mode can be made very small. We see that (16) agrees with (1) and that the electron stream provides a means for precisely measuring the mean squared electromotive force (including quantum effects) as given by (1), provided the transit angle is small.

CONCLUSION

We have studied the random changes of velocity of an electron stream interacting with a damped oscillator. The vacuum fluctuations are directly observable as noise in the electron stream if the circuit is at low temperature.

One might wonder why the vacuum (radiation) fields outside of the circuit do not also contribute to the zero point electron stream noise. The radiation fields outside do not contribute because a free electron cannot radiate. This is well known and results from the fact that the conditions of conservation of energy and conservation of momentum cannot be simultaneously satisfied. The electron can exchange energy with the fields of the circuit because during the interaction time the electron is not free. Its momentum is not precisely known during the interaction time, because its position is known to be localized in the interaction gap. The electron can therefore undergo spontaneous emission to the circuit, even if the circuit is in its lowest state. This loss of energy by the electron is a purely random process and therefore contributes to the noise. This is the origin of the zero point noise contribution in this case.

A factor which must be considered in doing a low-temperature noise experiment, is the noise (such as shot[6] noise), present in the beam before interaction with a resonator. The amount of this noise will depend on the manner in which the beam is prepared. The uncertainty principle does not preclude preparation of a low-current beam in which the electron velocities are closely grouped within a range substantially less than the velocity increment due to loss of a high-frequency microwave quantum. The vacuum fluctuation noise could be observed in the random changes of velocity of the electrons, after interaction. In any case, it is not essential that the noise in the electron beam before interaction, be less than the vacuum fluctuation noise. This is because the latter noise can be observed as an increment in the noise already present. Somewhat similar procedures are employed for temperature measurement by microwave radiometers.[7]

[6] W. Schottky, Ann. Physik **57**, 541 (1918).
[7] R. H. Dicke, Rev. Sci. Instr. **17**, 268 (1946).

Extension of Onsager's Theory of Reciprocal Relations. I

S. R. DE GROOT AND P. MAZUR
Institute for Theoretical Physics, University of Leyden, Leyden, Netherlands

(Received December 8, 1953)

A generalization of Onsager's proof for reciprocal relations between irreversible processes is developed in such a form that it can be used for vectorial and tensorial phenomena. This is achieved by means of an extension of the fluctuation formalism used by Onsager.

The method is applied to heat conduction, diffusion, viscosity, and cross-effects in anisotropic systems, and reciprocal relations are derived for these phenomena.

1. INTRODUCTION

ONSAGER[1] has established reciprocal relations between irreversible phenomena which he derived from the property of microscopic reversibility. These relations have found a wide field of application in the thermodynamics of irreversible processes.[2,3]

However, as Casimir[4] pointed out, Onsager's proof is strictly valid only for scalar phenomena. As a matter of fact Onsager assumes that the irreversible fluxes can be considered time derivates of thermodynamic state variables. This is correct for scalar processes such as chemical reactions and relaxation phenomena, but not for vectorial processes (such as heat conduction, diffusion and electrical conduction) and tensorial processes (such as viscous flow).

Casimir[4] has proposed a method to avoid this difficulty, on which he has elaborated for the case of heat conduction in anisotropic crystals. This method can briefly be described as follows. Equations are established for the regression of an infinite number of local temperature fluctuations. To this set of equations the ordinary Onsager scheme can be applied since the "fluxes" are now indeed time derivatives of state variables (here the local temperatures). In particular, reciprocal relations between corresponding irreversible processes (Onsager relations) are found. However, these relations are only an intermediate result, since the equations used are not identical with the ordinary macroscopic Fourier law of heat conduction. Only by means of a subsequent mathematical treatment was it possible to find the implications of the intermediate Onsager relations on the properties of the heat conduction tensor. In the derivation, constancy of the density of the system (i.e., absence of motion) was assumed.

The present authors[5] have applied a similar method to electrical and heat conduction in the presence of a magnetic field.

In this paper a different method is developed, which avoids the complications inherent in the approach outlined above. The method consists of a generalization of Onsager's original proof and enables one to derive directly reciprocal relations for irreversible processes of vectorial or tensorial nature.

2. FLUCTUATION THEORY

We consider an adiabatically insulated system. The thermodynamic state of the system is described by variables $A_1, A_2, \cdots, A_m$ which are even functions of the particle velocities, and variables $B_{m+1}, B_{m+2}, \cdots, B_n$ which are odd functions of the particle velocities[6] (the B-type variables, or their divergences, may be time derivatives of A-type variables: see also Machlup and Onsager[6] and Secs. 6 and 9). All variables are continuous functions of space and time coordinates. For convenience the system will be divided into a number of cells of volume V^μ, in which the thermodynamic variables may be considered as uniform (here μ numbers the cells). The deviations of the state variables from their equilibrium values are denoted by

$$\alpha_i^\mu = A_i^\mu - (A_i^\mu)_{\text{equ}}, \quad (i=1, 2, \cdots, m) \tag{1}$$

$$\beta_k^\mu = B_k^\mu - (B_k^\mu)_{\text{equ}}, \quad (k=m+1, m+2, \cdots, n). \tag{2}$$

The deviation of the entropy of the system from its equilibrium value is given by the quadratic form

$$\Delta S = -\tfrac{1}{2}\sum_{\mu,\nu} V^\mu V^\nu \Big(\sum_{i,j} g_{ij}{}^{\mu\nu}\alpha_i^\mu \beta_j^\nu + \sum_{k,l} h_{kl}{}^{\mu\nu}\beta_k^\mu \beta_l^\nu\Big). \tag{3}$$

Since the entropy is an even function of the particle velocities no cross-terms between α and β variables appear in the expression (3). The probability distribution for the α_i^μ and the β_k^μ is expressed by

$$P \prod_{\mu,i,k} d\alpha_i^\mu d\beta_k^\mu = \exp(\Delta S/k) \prod_{\mu,i,k} d\alpha_i^\mu d\beta_k^\mu \Big/ \int\cdots\int \exp(\Delta S/k) \prod_{\mu,i,k} d\alpha_i^\mu d\beta_k^\mu. \tag{4}$$

The following linear combinations of parameters are introduced:

$$X_i^\mu = (V^\mu)^{-1}\partial\Delta S/\partial\alpha_i^\mu = +\sum_\nu V^\nu \sum_j g_{ij}{}^{\mu\nu}\alpha_i^\nu, \tag{5}$$

$$Y_k^\mu = (V^\mu)^{-1}\partial\Delta S/\partial\beta_k^\mu = -\sum_\nu V^\nu \sum_l h_{kl}{}^{\mu\nu}\beta_l^\nu. \tag{6}$$

[1] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).
[2] I. Prigogine, *Etude thermodynamique des phénomènes irréversibles* (Editions Desoer, Liège, 1947).
[3] S. R. de Groot, *Thermodynamics of Irreversible Processes* (North-Holland Publishing Company, Amsterdam and Interscience Publishers, New York, 1951).
[4] H. B. G. Casimir, Revs. Modern Phys. **17**, 343 (1945).
[5] P. Mazur and S. R. de Groot, Physica **19**, 961, 1953.
[6] S. Machlup and L. Onsager, Phys. Rev. **91**, 1512 (1953).

A number of averages, which will be used later on, can easily be found with the help of (4), (5), and (6):

$$\langle \alpha_i^{\mu}(t) X_j^{\nu}(t) \rangle_{M} = -k\delta_{ij}\delta_{\mu\nu}/V^{\mu}, \tag{7}$$

$$\langle \beta_k^{\mu}(t) Y_l^{\nu}(t) \rangle_{M} = -k\delta_{kl}\delta_{\mu\nu}/V^{\mu}. \tag{8}$$

Average values of products of α- and β-type variables always vanish.

Passing to the limit of continuous variables, the indices μ and ν of (7) and (8), which indicate the cells, can be replaced by $\mathbf{r}$ and $\mathbf{r}'$, which denote space coordinates, whereas the last two factors of (7) and (8) combine into a Heaviside-Dirac δ function. Consequently (7) and (8) become

$$\langle \alpha_i(\mathbf{r},t) X_j(\mathbf{r}',t) \rangle_{M} = -k\delta_{ij}\delta(\mathbf{r}-\mathbf{r}'), \tag{9}$$

$$\langle \beta_k(\mathbf{r},t) Y_l(\mathbf{r}',t) \rangle_{M} = -k\delta_{kl}\delta(\mathbf{r}-\mathbf{r}'). \tag{10}$$

From (9) and (10) one can immediately derive the formulas

$$\langle \alpha_i(\mathbf{r},t)\Omega(\mathbf{r}') X_j(\mathbf{r}',t) \rangle_{M} = -k\delta_{ij}\Omega(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}'), \tag{11}$$

$$\langle \beta_k(\mathbf{r},t)\Omega(\mathbf{r}') Y_l(\mathbf{r}',t) \rangle_{M} = -k\delta_{kl}\Omega(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}'), \tag{12}$$

where $\Omega(\mathbf{r})$ is a differential operator of the general form

$$\Omega(\mathbf{r}) = \sum_{p,q,s} a_{p,q,s}(\mathbf{r}) \partial^{p+q+s} / \partial x_1^{p} \partial x_2^{q} \partial x_3^{s}. \tag{13}$$

The coefficients $a_{p,q,s}$ are independent of the state variables α and β; the cartesian coordinates are denoted by x_1, x_2 and x_3.

The formulas (11) and (12) will serve as an extremely useful basis for the straightforward derivation of the reciprocal relations amongst phenomenological coefficients.

Remark

It may be noted that with (5) and (6) the time derivative of (3) (entropy production per unit time) can be written as

$$d(\Delta S)/dt = \sum_{\mu} V^{\mu} \left(\sum_i \dot{\alpha}_i^{\mu} X_i^{\mu} + \sum_k \dot{\beta}_k^{\mu} Y_k^{\mu} \right), \tag{14}$$

or

$$d(\Delta S)/dt = \int \left\{ \sum_i \dot{\alpha}_i(\mathbf{r}) X_i(\mathbf{r}) + \sum_k \dot{\beta}_k(\mathbf{r}) Y_k(\mathbf{r}) \right\} d\mathbf{r}, \tag{15}$$

in the limit of infinitely small cells.

3. MICROSCOPIC REVERSIBILITY

As a result of the property of time reversal invariance of the equations of motion for individual particles, the α and β variables[3,4] satisfy the following relations:

$$\langle \alpha_i(\mathbf{r},t)\alpha_j(\mathbf{r}',t+\tau) \rangle_{M} = \langle \alpha_j(\mathbf{r}',t)\alpha_i(\mathbf{r},t+\tau) \rangle_{M}, \tag{16}$$

$$\langle \alpha_i(\mathbf{r},t)\beta_k(\mathbf{r}',t+\tau) \rangle_{M} = -\langle \beta_k(\mathbf{r}',t)\alpha_i(\mathbf{r},t+\tau) \rangle_{M}, \tag{17}$$

$$\langle \beta_k(\mathbf{r},t)\beta_l(\mathbf{r}',t+\tau) \rangle_{M} = \langle \beta_l(\mathbf{r}',t)\beta_k(\mathbf{r},t+\tau) \rangle_{M}. \tag{18}$$

From these three relations it follows that

$$\langle \alpha_i(\mathbf{r},t)(\partial/\partial t)\alpha_j(\mathbf{r}',t) \rangle_{M} = \langle \alpha_j(\mathbf{r}',t)(\partial/\partial t)\alpha_i(\mathbf{r},t) \rangle_{M}, \tag{19}$$

$$\langle \alpha_i(\mathbf{r},t)(\partial/\partial t)\beta_k(\mathbf{r}',t) \rangle_{M} = -\langle \beta_k(\mathbf{r}',t)(\partial/\partial t)\alpha_i(\mathbf{r},t) \rangle_{M}, \tag{20}$$

$$\langle \beta_k(\mathbf{r},t)(\partial/\partial t)\beta_l(\mathbf{r}',t) \rangle_{M} = \langle \beta_l(\mathbf{r}',t)(\partial/\partial t)\beta_k(\mathbf{r},t) \rangle_{M}. \tag{21}$$

Of course the time derivative in the last three relations should be interpreted on a microscope scale as a difference quotient;[4,5] however, for all practical macroscopic purposes this quotient can be considered as a real derivative.

For simplicity's sake we have not taken into consideration the influence of a magnetic field[5] or a rotation of the system on relations (16)–(18).

4. ENTROPY PRODUCTION IN LOCAL FORM

In order to apply the results derived so far to the actual physical processes, the macroscopic thermodynamical theory[2,3] of these phenomena must first be recalled (in this section and the following). Let us consider a system of n components in which the phenomena of heat conduction, diffusion and viscous flow and their cross-effects may occur. The system may also be subject to external conservative forces. The entropy production can be calculated from the fundamental laws of macroscopic physics. In the case under consideration these laws are:

The Mass Laws[7]

$$\rho dc_k/dt = -\operatorname{div}\mathbf{J}_k, \quad (k=1, 2, \cdots, n), \tag{22}$$

where ρ is the density, c_k the mass fraction of component k ($\sum_k c_k = 1$), and $\mathbf{J}_k$ the diffusion flow of k with respect to the center of mass motion ($\sum_k \mathbf{J}_k = 0$). Furthermore we have

$$\rho dv/dt = \operatorname{div}\mathbf{v}, \tag{23}$$

with $v = \rho^{-1}$, the specific volume, and $\mathbf{v}$ the barycentric velocity. In both Eqs. (22) and (23) d/dt stands for the substantial time derivative with respect to the center of mass motion

$$d/dt = \partial/\partial t + \mathbf{v}\cdot\operatorname{grad}. \tag{24}$$

With (24) the formulas (22) and (23) can alternatively be written as

$$\partial\rho_k/\partial t = -\operatorname{div}(\rho_k\mathbf{v}) - \operatorname{div}\mathbf{J}_k, \tag{25}$$

$$\partial\rho/\partial t = -\operatorname{div}(\rho\mathbf{v}), \tag{26}$$

where $\rho_k = c_k\rho$ is the density of component k.

The Momentum Law

$$\rho d\mathbf{v}/dt = -\operatorname{Div}\boldsymbol{P} + \sum_k \rho_k \mathbf{F}_k. \tag{27}$$

Here $\mathbf{F}_k$ is the external conservative force per unit

[7] For vector notation used here, see the Appendix.

mass acting on component k:

$$\mathbf{F}_k = -\text{grad}w_k; \quad \text{with} \quad \partial w_k/\partial t = 0, \tag{28}$$

where w_k is the potential energy per unit mass of component k. The pressure tensor $\boldsymbol{P}$ can be split into two parts

$$\boldsymbol{P} = p\boldsymbol{\delta} + \boldsymbol{\Pi}, \tag{29}$$

where $\boldsymbol{\delta}$ is the unit tensor, p the hydrostatic pressure, and $\boldsymbol{\Pi}$ the viscous pressure tensor.

With (24) and (26) one can write (27) alternatively as

$$\partial \mathbf{g}/\partial t = -\text{Div}(\mathbf{vg} + \boldsymbol{P}) + \textstyle\sum_k \rho_k \mathbf{F}_k, \tag{30}$$

where $\mathbf{g} = \rho\mathbf{v}$ is the momentum density.

The Energy Law

$$\rho de/dt = -\text{div}\mathbf{J}_e, \tag{31}$$

where e is the specific total energy and $\mathbf{J}_e$ the energy flow. These are given by

$$e = u + \tfrac{1}{2}\mathbf{v}^2 + \textstyle\sum_k c_k w_k, \tag{32}$$

$$\mathbf{J}_e = \mathbf{J}_q + \mathbf{v}\cdot\boldsymbol{P} + \textstyle\sum_k w_k \mathbf{J}_k, \tag{33}$$

with u the specific internal energy and $\frac{1}{2}\mathbf{v}^2$ the specific kinetic energy. Relation (33) defines the heat flow $\mathbf{J}_q$.

The first law of thermodynamics is found from (31) with (23) and (27)

$$\rho du/dt = -\text{div}\mathbf{J}_q - \boldsymbol{P}:\text{Grad}\mathbf{v} + \textstyle\sum_k \mathbf{F}_k\cdot\mathbf{J}_k. \tag{34}$$

Introducing the internal energy per unit volume $u_v = \rho u$ and applying (26), Eq. (34) can be written as

$$\partial u_v/\partial t = -\text{div}(u_v\mathbf{v} + \mathbf{J}_q) - \boldsymbol{P}:\text{Grad}\mathbf{v} - \textstyle\sum_k \mathbf{F}_k\cdot\mathbf{J}_k. \tag{35}$$

Gibbs' Law

$$Tds/dt = du/dt + pdv/dt - \textstyle\sum_k \mu_k dc_k/dt, \tag{36}$$

where T is the temperature, s the specific entropy, p the hydrostatic pressure (we consider fluids, where the elastic stress tensor reduces to a hydrostatic pressure only), and μ_k the chemical potential per unit mass.

Entropy Balance

With the help of (22), (23), and (34), Eq. (36) can be brought into the form of a balance equation

$$\rho ds/dt = -\text{div}\mathbf{J}_s + \sigma, \tag{37}$$

where $\mathbf{J}_s$ is the entropy flow,

$$\mathbf{J}_s = (\mathbf{J}_q - \textstyle\sum_k \mu_k\mathbf{J}_k)/T, \tag{38}$$

and where the entropy production σ follows from

$$T\sigma = -\mathbf{J}_q\cdot(\text{grad}T)/T + \textstyle\sum_k \mathbf{J}_k\cdot\{\mathbf{F}_k - T\,\text{grad}(\mu_k/T)\} - \boldsymbol{\Pi}:\text{Grad}\mathbf{v} \geqslant 0. \tag{39}$$

The three terms on the right-hand side are the result of entropy production caused by heat conduction, diffusion, and viscous flow, respectively.

Another useful form of $T\sigma$, which follows from (39) when the definition (38) of the entropy flow is used, is

$$T\sigma = -\mathbf{J}_s\cdot\text{grad}T - \textstyle\sum_k \mathbf{J}_k\cdot\text{grad}\bar{\mu}_k - \boldsymbol{\Pi}:\text{Grad}\mathbf{v} \geqslant 0, \tag{40}$$

where the quantity

$$\bar{\mu}_k \equiv \mu_k + w_k \tag{41}$$

has been introduced.

Just as (35) follows from (34), one can write instead (37),

$$\partial s_v/\partial t = -\text{div}(s_v\mathbf{v} + \mathbf{J}_s) + \sigma, \tag{42}$$

where $s_v = \rho s$ is the entropy per unit volume.

5. THE PHENOMENOLOGICAL EQUATIONS

The entropy production (39) or (40) is a sum of products of "fluxes" and so-called "forces" (or "affinities"). Linear relationships between these quantities can be established: these are called the phenomenological equations. Choosing the quantities occurring in (40) one can write

$$\mathbf{J}_s = -\boldsymbol{L}_{ss}\cdot\text{grad}T - \sum_{k=1}^{n} \boldsymbol{L}_{sk}\cdot\text{grad}\bar{\mu}_k, \tag{43}$$

$$\mathbf{J}_k = -\boldsymbol{L}_{ks}\cdot\text{grad}T - \sum_{k'=1}^{n} \boldsymbol{L}_{kk'}\cdot\text{grad}\bar{\mu}_{k'}, \quad (k = 1, 2, \cdots, n) \tag{44}$$

$$\Pi_{ij} = -\sum_{m,n=1}^{3} L_{ij,mn}(\text{Grad}\mathbf{v})_{mn}, \quad (i, j = 1, 2, 3) \tag{45}$$

where i, j, m and n indicate cartesian coordinates.

We are not concerned here with cross-effects between vectorial and tensorial quantities, although such effects might exist in anisotropic media. The present formalism could of course be easily extended to such a case. (In isotropic media these cross-effects do not exist according to the so-called Curie principle.)

The quantities $\boldsymbol{L}_{ss}$, $\boldsymbol{L}_{sk}$, $\boldsymbol{L}_{ks}$ and $\boldsymbol{L}_{kk'}$ (k, $k' = 1, 2, \cdots, n$) are tensors (of the second order). Certain relationships between these phenomenological coefficients exist, since not all fluxes and forces are independent. From $\sum_k \mathbf{J}_k = 0$ it follows that[3]

$$\sum_{k=1}^{n} \boldsymbol{L}_{sk} = 0, \quad \sum_{k=1}^{n} \boldsymbol{L}_{k's} = 0, \quad (k' = 1, 2, \cdots, n); \tag{46}$$

$$\sum_{k=1}^{n} \boldsymbol{L}_{ks} = 0, \quad \sum_{k=1}^{n} \boldsymbol{L}_{kk'} = 0. \quad (k' = 1, 2, \cdots, n). \tag{47}$$

Thus (43) and (44) can be rewritten with independent

fluxes and forces:

$$\mathbf{J}_s = -L_{ss}\cdot\mathrm{grad}T - \sum_{k=1}^{n-1} L_{sk}\cdot\mathrm{grad}(\bar{\mu}_k - \bar{\mu}_n), \tag{48}$$

$$\mathbf{J}_k = -L_{ks}\cdot\mathrm{grad}T - \sum_{k'=1}^{n-1} L_{kk'}\cdot\mathrm{grad}(\bar{\mu}_{k'} - \bar{\mu}_n). \quad (k=1, 2, \cdots, n-1). \tag{49}$$

In (48) and (49) we are left with n^2 phenomenological tensors (that is $9n^2$ phenomenological coefficients).

Let us now discuss the phenomenological equations (45). In the most general case the viscous pressure tensor $\mathbf{\Pi}$ has 9 independent components. We shall be concerned here, however, with the case usually considered of a symmetric tensor $\mathbf{\Pi}$, which has six independent components. Then in (45) only $(\mathrm{Grad}\mathbf{v})^s$, i.e., the symmetric part of (Grad$\mathbf{v}$) remains. Rewriting (45) for this case, we have

$$\Pi_{(ij)} = -\sum_{m,n=1}^{3} L_{(ij)(mn)}(\mathrm{Grad}\mathbf{v})_{(mn)}, \tag{50}$$

where the notations (ij) and (mn) indicate the symmetrical character in the indices between brackets. Consequently we are left with 36 phenomenological coefficients $L_{(ij)(mn)}$, instead of the 81 coefficients $L_{ij,mn}$ from (45). [When (Grad$\mathbf{v}$) and $\mathbf{\Pi}$ are brought to principle axes together, only 9 phenomenological coefficients are left.]

6. ENTROPY PRODUCTION IN TERMS OF FLUCTUATIONS

In order to apply the theory outlined in Secs. 2 and 3 to the phenomena described in the foregoing, one must proceed as follows: the change of an insulated system which in formula (3) was expressed in terms of arbitrary parameters α and β, must now be specified as a function of the variables which determine the actual physical state of the system considered in Secs. 4 and 5.

The change of entropy per unit time of the whole system of volume V is

$$dS/dt = \int_v (\partial s_v/\partial t)dV. \tag{51}$$

The local change of entropy $\partial s_v/\partial t$ can be expressed in terms of other state variables by means of the Gibbs' relation (36). This relation is valid along the center of gravity motion.[8] In the case under consideration (isolated system), this motion is a fluctuation. This means that in the Gibbs' relation terms in $\mathbf{v}$ [see formula (24)] are of higher order than the local time derivatives, and may therefore be neglected.[8] Consequently we can write

$$T\partial s/\partial t = \partial u/\partial t + p\partial v/\partial t - \sum_k \mu_k \partial c_k/\partial t, \tag{52}$$

or alternatively, since $\sum_k c_k\mu_k = u - Ts + pv$,

$$T\partial s_v/\partial t = \partial u_v/\partial t - \sum_k \mu_k \partial\rho_k/\partial t. \tag{53}$$

Introducing (53) into (51) we obtain

$$dS/dt = \int_v T^{-1}(\partial u_v/\partial t - \sum_k \mu_k \partial\rho_k/\partial t)dV, \tag{54}$$

which can also be written as

$$dS/dt = \int_v \{-(\Delta T/T_0)T^{-1}(\partial u_v/\partial t - \sum_k \mu_k \partial\rho_k/\partial t) + T_0^{-1}(\partial u_v/\partial t - \sum_k \mu_k \partial\rho_k/\partial t)\}dV, \tag{55}$$

where $\Delta T = T - T_0$, and where we take for T_0 the equilibrium temperature. The change of total energy E for the whole system follows from (32)

$$dE/dt = \int_V (\partial u_v/\partial t + \sum_k w_k \partial\rho_k/\partial t + \partial\tfrac{1}{2}\rho\mathbf{v}^2/\partial t)dV. \tag{56}$$

Introducing this expression and (53) into (55), one obtains

$$dS/dt = T_0^{-1}dE/dt - \int_V T_0^{-1}(\Delta T\partial s_v/\partial t + \sum_k \bar{\mu}_k \partial\rho_k/\partial t + \partial\tfrac{1}{2}\rho\mathbf{v}^2/\partial t)dV. \tag{57}$$

For an insulated system (constant E) (57) can finally be written as

$$T_0 d\Delta S/dt = -\int_V (\Delta T\partial\Delta s_v/\partial t + \sum_k \Delta\bar{\mu}_k \partial\Delta\rho_k/\partial t + \Delta\mathbf{v}\cdot\partial\Delta\mathbf{g}/\partial t)dV. \tag{58}$$

Here the symbol Δ indicates the difference of a quantity and its equilibrium value. Use has been made of the fact that *at equilibrium* all parameters are constant in time, T and μ_k are uniform, and $\mathbf{v}$ vanishes. Furthermore conservation of mass has been applied. In passing from (57) to (58) the difference between ρ and its equilibrium value ρ_0 is immaterial, since third order terms in (58) are neglected. For this reason, $\mathbf{g} \equiv \rho\mathbf{v}$ in (58) can be considered as $\rho_0\mathbf{v}$.

It may be remarked that the result (58) could also have been found from the volume integral of σ (40) under the same conditions and approximations. The derivation of (58) given here is, however, simpler.

It is clear that (58) is of the form (15), which is appropriate for the application of fluctuation theory (Sec. 2) and microscopic reversibility (Sec. 3). The variables $\Delta s_v(\mathbf{r})$, and $\Delta\rho_k(\mathbf{r})$ are of the α type, whereas the components of $\Delta\mathbf{g}(\mathbf{r})$ are β-type variables. (The variables ΔT, $\Delta\bar{\mu}_k$ and the components of $\Delta\mathbf{v}$ are the corresponding X- and Y-type variables.)

[8] I. Prigogine, Physica **15**, 272 (1949).

7. FLUCTUATIONS

To these variables we apply the relations (11) or (12). This gives

$$\langle \Delta s_v(\mathbf{r})\Omega(\mathbf{r}')\Delta T(\mathbf{r}')\rangle_{kl} = kT_0\Omega(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}'),$$

$$\langle \Delta s_v(\mathbf{r})\Omega(\mathbf{r}')\Delta\bar{\mu}_k(\mathbf{r}')\rangle_{kl} = 0, \tag{59}$$

$$\langle \Delta s_v(\mathbf{r})\Omega(\mathbf{r}')\Delta v_i(\mathbf{r}')\rangle_{kl} = 0,$$

$$\langle \Delta\rho_k(\mathbf{r})\Omega(\mathbf{r}')\Delta\bar{\mu}_{k'}(\mathbf{r}')\rangle_{kl} = kT_0\delta_{kk'}\Omega(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}'),$$

$$\langle \Delta\rho_k(\mathbf{r})\Omega(\mathbf{r}')\Delta T(\mathbf{r}')\rangle_{kl} = 0, \tag{60}$$

$$\langle \Delta\rho_k(\mathbf{r})\Omega(\mathbf{r}')\Delta v_i(\mathbf{r}')\rangle_{kl} = 0,$$

$$\langle \Delta g_i(\mathbf{r})\Omega(\mathbf{r}')\Delta v_j(\mathbf{r}')\rangle_{kl} = kT_0\delta_{ij}\Omega(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}'),$$

$$\langle \Delta g_i(\mathbf{r})\Omega(\mathbf{r}')\Delta T(\mathbf{r}')\rangle_{kl} = 0, \tag{61}$$

$$\langle \Delta g_i(\mathbf{r})\Omega(\mathbf{r}')\Delta\bar{\mu}_k(\mathbf{r}')\rangle_{kl} = 0,$$

where $i,j=1,2,3$ indicate Cartesian coordinates. Furthermore, there are no correlations between the β-type variables $\Delta\mathbf{v}$ and $\Delta\mathbf{g}$, and any combination of the α-type variables, because no cross-terms between α's and β's appear in expression (3).

8. RECIPROCAL RELATIONS FOR HEAT CONDUCTION, DIFFUSION, AND CROSS-EFFECTS

In order to derive reciprocal relations for heat conduction diffusion and their cross-effects the following expressions of microscopic reversibility, which are examples of relation (19), are needed

$$\langle \Delta s_v(\mathbf{r})(\partial/\partial t)\Delta s_v(\mathbf{r}')\rangle_{kl} = \langle \Delta s_v(\mathbf{r}')(\partial/\partial t)\Delta s_v(\mathbf{r})\rangle_{kl}, \tag{62}$$

$$\langle \Delta s_v(\mathbf{r})(\partial/\partial t)\Delta\rho_k(\mathbf{r}')\rangle_{kl} = \langle \Delta\rho_k(\mathbf{r}')(\partial/\partial t)\Delta s_v(\mathbf{r})\rangle_{kl},$$
$$(k=1,2,\cdots,n) \tag{63}$$

$$\langle \Delta\rho_k(\mathbf{r})(\partial/\partial t)\Delta\rho_{k'}(\mathbf{r}')\rangle_{kl} = \langle \Delta\rho_{k'}(\mathbf{r}')(\partial/\partial t)\Delta\rho_k(\mathbf{r})\rangle_{kl}.$$
$$(k,k'=1,2,\cdots,n). \tag{64}$$

Introducing into these relations the entropy balance (42), the mass laws (25), and the phenomenological equations (48) and (49), one obtains

$$\langle \Delta s_v[\mathrm{div}'\{-s_v'\Delta\mathbf{v}'+L_{sv}'\cdot\mathrm{grad}'\Delta T' + \sum_{k=1}^{n-1} L_{sk}'\cdot\mathrm{grad}'(\Delta\bar{\mu}_k'-\Delta\bar{\mu}_n')\}+\sigma']\rangle_{kl}$$
$$= \langle \Delta s_v'[\mathrm{div}\{-s_v\Delta\mathbf{v}+L_{sv}\cdot\mathrm{grad}\Delta T + \sum_{k=1}^{n-1} L_{sk}\cdot\mathrm{grad}(\Delta\bar{\mu}_k-\Delta\bar{\mu}_n)\}+\sigma]\rangle_{kl}, \tag{65}$$

$$\langle \Delta s_v[\mathrm{div}'\{-\rho_k'\Delta\mathbf{v}'+L_{ks}'\cdot\mathrm{grad}'\Delta T' + \sum_{k'=1}^{n-1} L_{kk'}'\cdot\mathrm{grad}'(\Delta\bar{\mu}_{k'}'-\Delta\bar{\mu}_n')\}]\rangle_{kl}$$
$$= \langle \Delta\rho_k'[\mathrm{div}\{-s_v\Delta\mathbf{v}+L_{sv}\cdot\mathrm{grad}\Delta T + \sum_{k'=1}^{n-1} L_{sk'}\cdot\mathrm{grad}(\Delta\bar{\mu}_{k'}-\Delta\bar{\mu}_n)\}+\sigma]\rangle_{kl}, \tag{66}$$

$$\langle \Delta\rho_k[\mathrm{div}'\{-\rho_{k'}'\Delta\mathbf{v}'+L_{k's}'\cdot\mathrm{grad}'\Delta T' + \sum_{k''=1}^{n-1} L_{k'k''}'\cdot\mathrm{grad}'(\Delta\bar{\mu}_{k''}'-\Delta\bar{\mu}_n')\}]\rangle_{kl}$$
$$= \langle \Delta\rho_{k'}'[\mathrm{div}\{-\rho_k\Delta\mathbf{v}+L_{ks}\cdot\mathrm{grad}\Delta T + \sum_{k''=1}^{n-1} L_{kk''}\cdot\mathrm{grad}(\Delta\bar{\mu}_{k''}-\Delta\bar{\mu}_n)\}]\rangle_{kl}, \tag{67}$$

where dashes indicate dependence on $\mathbf{v}'$. In these relations occur the quantities

$$s_v\Delta\mathbf{v} = s_v^0\Delta\mathbf{v}+\Delta s_v\Delta\mathbf{v}, \tag{68}$$

$$\rho_k\Delta\mathbf{v} = \rho_k^0\Delta\mathbf{v}+\Delta\rho_k\Delta\mathbf{v}, \tag{69}$$

where s_v^0 and ρ_k^0 are equilibrium values.

Applying (59) and (60) to (65), (66) and (67) one gets

$$\mathrm{div}'\{L_{ss}'\cdot\mathrm{grad}'\delta(\mathbf{r}-\mathbf{r}')\} = \mathrm{div}\{L_{ss}\cdot\mathrm{grad}\delta(\mathbf{r}-\mathbf{r}')\}, \tag{70}$$

$$\mathrm{div}'\{L_{ks}'\cdot\mathrm{grad}'\delta(\mathbf{r}-\mathbf{r}')\}$$
$$= \mathrm{div}\{\sum_{k'=1}^{n-1} L_{sk'}\cdot(\delta_{kk'}-\delta_{kn})\,\mathrm{grad}\delta(\mathbf{r}-\mathbf{r}')\}, \tag{71}$$

$$\mathrm{div}'\{\sum_{k''=1}^{n-1} L_{k'k''}'\cdot(\delta_{kk''}-\delta_{kn})\,\mathrm{grad}'\delta(\mathbf{r}-\mathbf{r}')\}$$
$$= \mathrm{div}\{\sum_{k''=1}^{n-1} L_{kk''}\cdot(\delta_{k'k''}-\delta_{k'n})\,\mathrm{grad}\delta(\mathbf{r}-\mathbf{r}')\}, \tag{72}$$

since averages of third order terms, e.g. $\langle \Delta\rho_k'\Delta s_v\Delta\mathbf{v}\rangle_{kl}$ or $\langle \Delta s_v\sigma'\rangle_{kl}$, vanish.

Eliminating the Kronecker δ's and the δ functions[3] from the preceding relations one finds

$$\mathrm{Div}L_{ss} = \mathrm{Div}L_{ss}^{\dagger}, \tag{73}$$

$$L_{sk}+L_{sk}^{\dagger} = L_{ks}+L_{ks}^{\dagger},$$
$$\mathrm{Div}L_{sk} = \mathrm{Div}L_{ks}^{\dagger}, \quad (k=1,2,\cdots,n-1) \tag{74}$$

$$L_{k'k}+L_{k'k}^{\dagger} = L_{kk'}+L_{kk'}^{\dagger},$$
$$\mathrm{Div}L_{k'k} = \mathrm{Div}L_{kk'}^{\dagger}, \quad (k,k'=1,2,\cdots,n-1), \tag{75}$$

where the symbol † indicates transposing of the Cartesian coordinates. With the convention of taking all phenomenological coefficients zero in empty space, and the fact that these coefficients do not depend on the shape of the sample it follows

$$L_{ss} = L_{ss}^{\dagger}, \tag{76}$$

$$L_{sk} = L_{ks}^{\dagger}, \quad (k=1,2,\cdots,n-1) \tag{77}$$

$$L_{kk'} = L_{k'k}^{\dagger}. \quad (k,k'=1,2,\cdots,n-1). \tag{78}$$

Taking into account (46) and (47), relations (77) and (78) are also valid for k, or k', or both, equal to n.

The formulas (76)–(78) are the reciprocal relations, which we wanted to derive for the phenomena of heat conduction, diffusion and cross-effects in anisotropic media.

The preceding theory could alternatively have been developed starting for instance from the form (39) of the entropy production, and corresponding phenomenological equations and sets of variables in Secs. 6 and 7.

Finally, we would like to make the following remark. It may at first sight seem surprising that relations (77) and (78) are found only for values of k and k' up to $n-1$, whereas in the starting point (63) and (64) they could also be equal to n. This can be understood because effectively one of the relations (63), and $2n-1$ of the relations (64) yield no information. From (64) for instance it is seen by summing over k, k' that one obtains

$$\langle\Delta\rho(\mathbf{r})(\partial/\partial t)\Delta\rho(\mathbf{r}')\rangle_{hr}=\langle\Delta\rho(\mathbf{r}')(\partial/\partial t)\Delta\rho(\mathbf{r})\rangle_{hr}. \quad (79)$$

Substituting (26) into this last relation this gives

$$\langle\Delta\rho\ \mathrm{div}'\Delta\mathbf{g}'\rangle_{hr}=\langle\Delta\rho'\ \mathrm{div}\Delta\mathbf{g}\rangle_{hr}. \quad (80)$$

According to the argument at the end of Sec. 7, both members of (80) vanish. Machlup and Onsager[4] have already discussed that correlations between α variables of the type (79) disappear when β variables, related to the α's in a way as expressed by (26), are necessary to describe the thermodynamic state of the system.

9. RECIPROCAL RELATIONS FOR VISCOSITY

Reciprocal relations for viscosity can be derived from the following expression for microscopic reversibility, which is an example of (21)

$$\langle\Delta g_i(\mathbf{r})(\partial/\partial t)\Delta g_j(\mathbf{r}')\rangle_{hr}=\langle\Delta g_j(\mathbf{r}')(\partial/\partial t)\Delta g_i(\mathbf{r})\rangle_{hr}. \quad (i,\ j=1,\ 2,\ 3). \quad (81)$$

Substituting (30) with (29) and the phenomenological equation (50) into (81) we find

$$\Big\langle\Delta g_i\Big[\sum_{l=1}^{3}\frac{\partial}{\partial x_l'}\Big\{-\Delta v_l'\Delta g_j'+\sum_{m,n=1}^{3}L'_{(lj)(mn)}\times\frac{1}{2}\Big(\frac{\partial\Delta v_n'}{\partial x_m'}+\frac{\partial\Delta v_m'}{\partial x_n'}\Big)\Big\}+\sum_{k=1}^{n}\rho_k'\Big(\frac{\partial\Delta\bar{\mu}_k'}{\partial x_j'}\Big)_T\Big]\Big\rangle_{hr}$$

$$=\Big\langle\Delta g_j'\Big[\sum_{l=1}^{3}\frac{\partial}{\partial x_l}\Big\{-\Delta v_l\Delta g_i+\sum_{m,n=1}^{3}L_{(li)(mn)}\times\frac{1}{2}\Big(\frac{\partial\Delta v_n}{\partial x_m}+\frac{\partial\Delta v_m}{\partial x_n}\Big)\Big\}+\sum_{k=1}^{n}\rho_k\Big(\frac{\partial\Delta\bar{\mu}_k}{\partial x_i}\Big)_T\Big]\Big\rangle_{hr}. \quad (82)$$

Here use has been made of

$$\sum_{k=1}^{n}\rho_k\mathbf{F}_k-\mathrm{grad}p=-\sum_{k=1}^{n}\rho_k\{\mathrm{grad}(w_k+\mu_k)\}_T=-\sum_{k=1}^{n}\rho_k(\mathrm{grad}\bar{\mu}_k)_T=-\sum_{k=1}^{n}\rho_k(\mathrm{grad}\Delta\bar{\mu}_k)_T, \quad (83)$$

in which the Gibbs-Duhem relation has been applied, and where the subscript T means constant temperature:

$$\sum_k\rho_k(\mathrm{grad}\Delta\bar{\mu}_k)_T=\sum_k\rho_k\,\mathrm{grad}\Delta\bar{\mu}_k-s_v\,\mathrm{grad}\Delta T. \quad (84)$$

Since averages of third order terms vanish and since no correlations exist between α- and β-type variables, we find from (82), with the help of (61)

$$\sum_{l,m,n}\frac{\partial}{\partial x_l'}\Big[L'_{(lj)(mn)}\frac{1}{2}\Big\{\frac{\partial}{\partial x_m'}\delta_{in}\delta(\mathbf{r}-\mathbf{r}')+\frac{\partial}{\partial x_n'}\delta_{im}\delta(\mathbf{r}-\mathbf{r}')\Big\}\Big]$$

$$=\sum_{l,m,n}\frac{\partial}{\partial x_l}\Big[L_{(li)(mn)}\frac{1}{2}\Big\{\frac{\partial}{\partial x_m}\delta_{jn}\delta(\mathbf{r}-\mathbf{r}')+\frac{\partial}{\partial x_n}\delta_{jm}\delta(\mathbf{r}-\mathbf{r}')\Big\}\Big]. \quad (85)$$

Eliminating the Kronecker δ's and the δ functions, we find the following results:

$$L_{(lj)(mi)}+L_{(mj)(li)}=L_{(li)(mj)}+L_{(mi)(lj)}, \quad (86)$$

$$\sum_m(\partial/\partial x_m)L_{(lj)(mi)}=\sum_m(\partial/\partial x_m)L_{(mi)(lj)}. \quad (87)$$

With the convention that the coefficients vanish in empty space this gives

$$L_{(lj)(mi)}=L_{(mi)(lj)}. \quad (i,\ j,\ l,\ m=1,\ 2,\ 3). \quad (88)$$

These are 15 reciprocal relations amongst the 36 viscosity coefficients. This leaves us with 21 independent coefficients.

It can be noted that in isotropic media one has two independent viscosity coefficients, but no reciprocal relations.

10. CONCLUDING REMARKS

In the preceding sections, reciprocal relations (76)–(78), (88) have been derived for vectorial and tensorial phenomena in continuous systems. In this way general results have been found for heat conduction, diffusion, viscosity and cross-effects in moving media. However, we did not include chemical reactions and other scalar phenomena into the considerations, since Onsager's formalism can be applied to these effects without any extension of the theory. Furthermore, only conservative external forces have been taken into account, which means that electromagnetic phenomena have not been dealt with. This will be done in a subsequent paper.

APPENDIX ON NOTATION

We use essentially the Milne system of tensor notation.[9] The exterior product of an ordered pair of

[9] L. Rosenfeld, *Theory of Electrons* (North-Holland Publishing Company, Amsterdam, 1951).

vectors $\mathbf{a}$, $\mathbf{b}$ is a tensor

$$\boldsymbol{T}=\mathbf{ab}, \text{ components: } T_{ik}=a_i b_k. \tag{89}$$

The divergence of a tensor and the gradient of a vector are written as

$$\mathrm{Div}\boldsymbol{T}, \text{ components: } (\mathrm{Div}\boldsymbol{T})_i=\sum_k(\partial/\partial x_k)T_{ki}, \tag{90}$$

$$\mathrm{Grad}\mathbf{a}, \text{ components: } (\mathrm{Grad}\mathbf{a})_{ik}=(\partial/\partial x_i)a_k. \tag{91}$$

The interior products between a tensor and a vector are denoted by

$$\boldsymbol{T}\cdot\mathbf{a}, \text{ components: } (\boldsymbol{T}\cdot\mathbf{a})_i=\sum_k T_{ik}a_k, \tag{92}$$

$$\mathbf{a}\cdot\boldsymbol{T}, \text{ components: } (\mathbf{a}\cdot\boldsymbol{T})_i=\sum_k a_k T_{ki}. \tag{93}$$

Finally,

$$\boldsymbol{T}:\boldsymbol{U}=\sum_{i,k} T_{ik}U_{ki} \tag{94}$$

is the interior product of two tensors, $\boldsymbol{T}$ and $\boldsymbol{U}$, contracted twice.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Extension of Onsager's Theory of Reciprocal Relations. II

P. MAZUR AND S. R. DE GROOT
Institute for Theoretical Physics, University of Leyden, Leyden, Netherlands
(Received December 7, 1953)

The extension of Onsager's theory of reciprocal relations between irreversible processes, developed previously by the authors, is applied to electric conduction in anisotropic crystals, and the symmetry properties of the conduction tensor are derived.

1. INTRODUCTION

IN a previous paper[1] the authors have extended Onsager's theory for reciprocal relations between irreversible processes in such a way that it could be directly applied to vectorial and tensorial phenomena (see I, Sec. 1). Results were obtained for heat conduction, diffusion, viscosity, and cross effects. It was assumed that external forces, taken into account, were conservative. Thus electromagnetic irreversible phenomena were not treated. It is the purpose of this paper to apply the general theory outlined in I, Secs. 2 and 3 to electromagnetic processes. We shall consider the case of electric conduction in anisotropic crystals at uniform temperature. In Secs. 2 and 3 the macroscopic phenomenological theory of electric conduction is developed. In Sec. 4 we then derive an expression for the entropy production in an energetically insulated crystal in terms of local fluctuations of state variables, as required for the application in Secs. 5–7 of the formalism from I, Secs. 2 and 3. In such a way the symmetry properties of the conduction tensor are derived.

2. ENTROPY PRODUCTION IN LOCAL FORM

Let us consider a system, consisting of a rigid ion lattice and of electrons, in an electromagnetic field. The entropy production can be calculated as follows:

Charge conservation is expressed by

$$\partial\rho_e/\partial t=-\mathrm{div}\mathbf{i}, \tag{1}$$

where ρ_e is the electrical charge density, and $\mathbf{i}$ the electric current.

Conservation of energy can be written as

$$\partial e_v/\partial t=-\mathrm{div}\mathbf{J}_e, \tag{2}$$

where e_v is the density of total energy and $\mathbf{J}_e$ the energy flow. These are given by

$$e_v=u_v+\tfrac{1}{2}(\mathbf{E}^2+\mathbf{B}^2), \tag{3}$$

$$\mathbf{J}_e=\mathbf{J}_q+c\mathbf{E}\times\mathbf{B}, \tag{4}$$

with u_v the density of internal energy, $\frac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)$ the density of electromagnetic energy ($\mathbf{E}$ is the electric and $\mathbf{B}$ the magnetic field), and $c\mathbf{E}\times\mathbf{B}$ the Poynting vector. Relation (4) defines the heat flow $\mathbf{J}_q$. Subtracting Poynting's theorem

$$\tfrac{1}{2}\partial(\mathbf{E}^2+\mathbf{B}^2)/\partial t=-\mathrm{div}c\mathbf{E}\times\mathbf{B}+\mathbf{i}\cdot\mathbf{E}, \tag{5}$$

from (2) one obtains the equation for the internal energy

$$\partial u_v/\partial t=-\mathrm{div}\mathbf{J}_q+\mathbf{i}\cdot\mathbf{E}. \tag{6}$$

Furthermore, we need the Gibbs' equation

$$Tds_v/dt=du_v/dt-\sum_k \mu_k d\rho_k/dt, \tag{7}$$

where T is the temperature, s_v the density of entropy, μ_k and ρ_k the chemical potential per unit mass and the density of component k (ions or electrons). The time derivatives in this equation are substantial derivatives with respect to the center of mass motion. Taking the velocity of the ion lattice zero we can neglect the center of mass motion, because the ions are heavy as compared

[1] S. R. de Groot and P. Mazur, preceding paper [Phys. Rev. 94, 218 (1954)], referred to in the following as I.

to the electrons. We have then instead of (7):

$$T\partial s_v/\partial t=\partial u_v/\partial t-(\mu/e)\partial\rho_e/\partial t, \tag{8}$$

where μ is the chemical potential and e the charge per unit mass of the electrons. We have taken into account the fact that the density of the ions is constant (ρ_e is the electrical charge density due to both electrons and ions).

In accordance with the approximation made above no equation of motion is needed, whilst no convection terms occur in (4).

From (1), (6) and (8) the entropy balance equation can be derived

$$\partial s_v/\partial t=-\mathrm{div}\mathbf{J}_s+\sigma, \tag{9}$$

with the entropy flow

$$\mathbf{J}_s=(\mathbf{J}_q-\mathbf{i}\mu/e)/T, \tag{10}$$

and the entropy source σ from

$$\begin{aligned}T\sigma&=-\mathbf{J}_q\cdot(\mathrm{grad}T)/T+\mathbf{i}\cdot\{\mathbf{E}-T\,\mathrm{grad}(\mu/Te)\}\\&=-\mathbf{J}_s\cdot\mathrm{grad}T+\mathbf{i}\cdot\{\mathbf{E}-\mathrm{grad}(\mu/e)\}\geqq 0.\end{aligned} \tag{11}$$

At uniform temperature (11) reduces to

$$T\sigma=\mathbf{i}\cdot\{\mathbf{E}-\mathrm{grad}(\mu/e)\}\geqq 0. \tag{12}$$

Expressing $\mathbf{E}$ in the electromagnetic potentials φ and $\mathbf{A}$

$$E=-\mathrm{grad}\varphi-c^{-1}\partial\mathbf{A}/\partial t, \tag{13}$$

(12) becomes

$$T\sigma=-\mathbf{i}\cdot\{\mathrm{grad}(\bar{\mu}/e)+c^{-1}\partial\mathbf{A}/\partial t\}\geqq 0, \tag{14}$$

where

$$\bar{\mu}=\mu+e\varphi \tag{15}$$

is called the electrochemical potential of the electrons.

3. THE PHENOMENOLOGICAL EQUATION

With the flux and the force occurring in (14) we can establish the linear relationship

$$\mathbf{i}=-\boldsymbol{L}(\mathbf{B})\cdot\{\mathrm{grad}(\bar{\mu}/e)+c^{-1}\partial\mathbf{A}/\partial t\}, \tag{16}$$

which is Ohm's law. Here $\boldsymbol{L}(\mathbf{B})$ is the electrical conductivity tensor which can be a function of the magnetic field strength $\mathbf{B}$. This equation can alternatively be written as

$$c^{-1}\partial\mathbf{A}/\partial t=-\mathrm{grad}(\bar{\mu}/e)-\boldsymbol{R}(\mathbf{B})\cdot\mathbf{i}, \tag{17}$$

where $\boldsymbol{R}(\mathbf{B})\equiv\boldsymbol{L}^{-1}(\mathbf{B})$ is the resistivity tensor.

4. ENTROPY PRODUCTION IN TERMS OF FLUCTUATIONS

The entropy production $d_\nu S/dt$ of the whole system, kept at uniform temperature, can be found by integrating (14) over the volume of the system,

$$\begin{aligned}d_\nu S/dt&=-T^{-1}\int_V\mathbf{i}\cdot\{\mathrm{grad}(\bar{\mu}/e)+c^{-1}\partial\mathbf{A}/\partial t\}dV\\&=-T^{-1}\int_V\{(\bar{\mu}/e)\partial\rho_e/\partial t+\mathbf{i}\cdot c^{-1}\partial\mathbf{A}/\partial t\}dV\\&\quad-T^{-1}\int_\Omega(\bar{\mu}\mathbf{i}/e)\cdot d\boldsymbol{\Omega},\end{aligned} \tag{18}$$

where (1) has been used and Gauss' theorem applied. The last term, which is an integral over the surface $\boldsymbol{\Omega}$, vanishes when the system is materially insulated ($\mathbf{i}$ vanishes at the surface).

If furthermore the system is also insulated for heat, it follows from the volume integral of (9) with (10), and from (18) that the change in the entropy of the whole system with time is

$$dS/dt=-T^{-1}\int_V\{(\bar{\mu}/e)\partial\rho_e/\partial t+\mathbf{i}\cdot c^{-1}\partial\mathbf{A}/\partial t\}dV, \tag{19}$$

or, taking into account that at equilibrium $\bar{\mu}$ is uniform, $\mathbf{i}$ vanishes and all variables are constant in time,

$$\begin{aligned}d\Delta S/dt=-T^{-1}\int_V\{&(\Delta\bar{\mu}/e)\partial\Delta\rho_e/\partial t\\&+\Delta\mathbf{i}\cdot c^{-1}\partial\Delta\mathbf{A}/\partial t\}dV,\end{aligned} \tag{20}$$

where conservation of charge has been applied, and where the symbol Δ indicates the difference of a quantity and its equilibrium value. We have obtained the change of entropy as a function of fluctuations of local state variables. It has the form (I, 15) required for the application of the formalism derived in I, Secs. 2 and 3. However, this formalism is based on the consideration of an *energetically* insulated system. It is therefore necessary to investigate whether such a condition leaves expression (20) unmodified. For this purpose let us write down the law of conservation of energy for the whole system which follows from (2), (3), and (4):

$$\begin{aligned}(d/dt)\int_V u_\nu dV+(d/dt)\int_V\tfrac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)dV\\=-\int_\Omega\mathbf{J}_q\cdot d\boldsymbol{\Omega}-c\int_\Omega(\mathbf{E}\times\mathbf{B})\cdot d\boldsymbol{\Omega}.\end{aligned} \tag{21}$$

Energetic insulation means that no energy fluxes enter through the limiting surface, i.e., both terms on the right-hand side of (21) vanish. The fact that the heat flow $\mathbf{J}_q$ must vanish at the boundary has been used already as a sufficient condition to obtain formula (20). On the other hand electromagnetic insulation of the system, i.e., vanishing of the last term of (21), is impossible since varying charges and currents give rise to energy changes in free space. Therefore, energetic

insulation must imply here the constancy of a quantity

$$\int_V u_v dV + \int_\infty \tfrac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)dV, \tag{22}$$

where the first integral is taken over the material system, and the second is to be extended over the whole of space. Of course the material system must contain all relevant sources of electromagnetic fields. If now the fields vanish sufficiently rapidly at infinity

$$c\int_\infty (\mathbf{E}\times\mathbf{B})\cdot d\mathbf{\Omega}$$

vanishes and constancy of (22) is achieved.

5. FLUCTUATIONS

In formula (20) we have the variables $\Delta\rho_e(\mathbf{r})$ of the α type, whereas the components of $\Delta\mathbf{A}(\mathbf{r})$ are of β-type variables (see I, Sec. 2). They are continuous function of space coordinates $\mathbf{r}$. The quantities $\Delta(\bar{\mu}/e)$, and the components of $\Delta\mathbf{i}$ are the corresponding X- and Y-type variables. The variables $\Delta\mathbf{i}$ and $\Delta\rho_e$ are connected by Eq. (1).

The relation (I, 11) and (I, 12) read for these variables

$$\left.\begin{aligned}\langle\Delta\rho_e(\mathbf{r})\Omega(\mathbf{r}')\Delta\bar{\mu}(\mathbf{r}')/e\rangle_{\mathcal{M}} &= kT\Omega(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}'),\\ \langle\Delta\rho_e(\mathbf{r})\Omega(\mathbf{r}')\Delta i_i(\mathbf{r}')\rangle_{\mathcal{M}} &= 0, \quad (i=1,2,3)\end{aligned}\right\} \tag{23}$$

$$\left.\begin{aligned}\langle\Delta A_i(\mathbf{r})\Omega(\mathbf{r}')\Delta i_j(\mathbf{r}')\rangle_{\mathcal{M}} &= kT\Omega(\mathbf{r}')\delta_{ij}\delta(\mathbf{r}-\mathbf{r}'),\\ &\qquad (i,j=1,2,3)\\ \langle\Delta A_i(\mathbf{r})\Omega(\mathbf{r}')\Delta\bar{\mu}(\mathbf{r}')/e\rangle_{\mathcal{M}} &= 0, \quad (i=1,2,3)\end{aligned}\right\} \tag{24}$$

where i and j indicate the Cartesian coordinates.

6. MICROSCOPIC REVERSIBILITY

Microscopic reversibility can be expressed as follows

$$\langle\Delta\rho_e(\mathbf{r})(\partial/\partial t)\Delta\rho_e(\mathbf{r}')\rangle_{\mathcal{M}}\{\mathbf{B},\mathbf{B}'\} = \langle\Delta\rho_e(\mathbf{r}')(\partial/\partial t)\Delta\rho_e(\mathbf{r})\rangle_{\mathcal{M}}\{-\mathbf{B},-\mathbf{B}'\}, \tag{25}$$

$$\langle\Delta\rho_e(\mathbf{r})(\partial/\partial t)\Delta A_i(\mathbf{r}')\rangle_{\mathcal{M}}\{\mathbf{B},\mathbf{B}'\} = -\langle\Delta A_i(\mathbf{r}')(\partial/\partial t)\Delta\rho_e(\mathbf{r})\rangle_{\mathcal{M}}\{-\mathbf{B},-\mathbf{B}'\}, \tag{26}$$

$$\langle\Delta A_i(\mathbf{r})(\partial/\partial t)\Delta A_j(\mathbf{r}')\rangle_{\mathcal{M}}\{\mathbf{B},\mathbf{B}'\} = \langle\Delta A_j(\mathbf{r}')(\partial/\partial t)\Delta A_i(\mathbf{r})\rangle_{\mathcal{M}}\{-\mathbf{B},-\mathbf{B}'\}, \tag{27}$$

which are examples of (I, 19), (I, 20), and (I, 21) respectively. With time reversal, i.e., all particles retracing their path, the magnetic field reverses its direction everywhere. In formulas (25)–(27), $\mathbf{B}$ and $\mathbf{B}'$ indicate the magnetic field strengths at the positions $\mathbf{r}$ and $\mathbf{r}'$, where the averages are performed.

7. RECIPROCAL RELATIONS FOR ELECTRIC CONDUCTION

In order to derive reciprocal relations for electric conduction in anisotropic crystals, we insert the law of conservation of charge (1) and the phenomenological equation (17) into the expressions for the microscopic reversibility (25)–(27),

$$\langle\Delta\rho_e\,\mathrm{div}'\Delta\mathbf{i}'\rangle_{\mathcal{M}} = \langle\Delta\rho_e'\,\mathrm{div}\Delta\mathbf{i}\rangle_{\mathcal{M}}, \tag{28}$$

$$\langle\Delta\rho_e\{(\partial/\partial x_i')\Delta\bar{\mu}/e+\textstyle\sum_j R_{ij}'(\mathbf{B}')\Delta i_j'\}\rangle_{\mathcal{M}} = -\langle\Delta A_i'\textstyle\sum_j(\partial/\partial x_j)\Delta i_j\rangle_{\mathcal{M}}, \tag{29}$$

$$\langle\Delta A_i\{(\partial/\partial x_j')\Delta\bar{\mu}/e+\textstyle\sum_k R_{jk}'(\mathbf{B}')\Delta i_k'\}\rangle_{\mathcal{M}} = \langle\Delta A_j'\{(\partial/\partial x_i)\Delta\bar{\mu}/e+\textstyle\sum_k R_{ik}(-\mathbf{B})\Delta i_k\}\rangle_{\mathcal{M}}. \tag{30}$$

With the (23) and (24) we find that both members of (28) vanish, whereas (29) and (30) become

$$(\partial/\partial x_i')\delta(\mathbf{r}-\mathbf{r}') = -\textstyle\sum_j\delta_{ij}(\partial/\partial x_j)\delta(\mathbf{r}-\mathbf{r}'), \tag{31}$$

$$\textstyle\sum_k R_{jk}'(\mathbf{B}')\delta_{ik}\delta(\mathbf{r}-\mathbf{r}') = \sum_k R_{ik}(-\mathbf{B})\delta_{jk}\delta(\mathbf{r}-\mathbf{r}'). \tag{32}$$

Eliminating the Kronecker δ's and the δ functions, (31) gives an identity whereas (32) yields the result

$$R_{ji}(\mathbf{B}) = R_{ij}(-\mathbf{B}), \tag{33}$$

or, in matrix notation

$$R(\mathbf{B}) = R^{\dagger}(-\mathbf{B}), \tag{34}$$

where the sign † indicates the transpose matrix. Formula (34) is the desired result for the symmetry properties of the electrical resistivity tensor.

We may remark that (25) did not lead to any result, since both terms became identically equal to zero [see (28)]. Such a behavior for a correlation between α-type variables, when β-type variables, related to the α's, are taken into consideration for the description of the system, was also found in I, Sec. 8 and was first noted by Machlup and Onsager.[2] We have also seen that (26) leads to no physical result [see (29) and (31)] since it is an identity. This means that a correlation between α variables and a related β variable is identically equal (but with opposite sign!) to the correlation between this β and this α. Or in other words, the property of microscopic reversibility is identically satisfied in this case. The complete result (34) is obtained from relation (27) alone.

The authors are indebted to Dr. R. Fieschi for some useful comments.

[2] S. Machlup and L. Onsager, Phys. Rev. **91**, 1512 (1953).

Some Aspects of Breakdown Streamers*

LEONARD B. LOEB
Department of Physics, University of California, Berkeley, California
(Received December 7, 1953)

Photomultiplier and oscillographic study of positive breakdown streamers yields values of current and distance of tip advance as a function of time. From these the relatively constant number n of excess positive ions left behind by the streamer per cm advance and the tip velocity v_t are obtained. These are related to the linear electron density n_e per cm of streamer channel plasma and the electron drift velocity v_e in the channel by $n_e v_e = n v_t$, with $v_t > v_e$. Since v_e is related to the gradient X_e in the channel by well-known drift velocity data, it is possible to limit the range of values of n_e and v_e. The limiting conditions are that $n_e > n$, $\int_0^s X_e dx < V_s$, where V_s is the applied potential, and finally that X_e must be sufficiently large to maintain the channel conducting despite dissociative recombination and dissociative attachment losses. In longer channels heating cannot be neglected. The dense positive space charge left behind by the electron current in the ionized channel of radius R causes a radial expansion of the excess space charge to a radius R' by incoming electron avalanches leading to a broad transient luminosity for some distance behind the tip. The tip and the conducting channel taking the return stroke, however, retain their small radius R.

THE streamer mechanism of breakdown leading to filamentary sparks initially proposed by H. Raether and independently by J. M. Meek and the writer, has been generally fairly well established through the observations of Lawrence and Dunnington,[1] by Raether[2] for uniform field geometry and by the writer and his students in corona studies, and by Allibone and Meek[3] and Meek and Saxe[4] in point to plane impulse breakdown. The studies of Fisher[5] and his associates and Bandel[6] in the writer's laboratory have shown that while very near threshold the breakdown starts as a low-order Townsend discharge with photoelectric liberation at the cathode, the eventual filamentary spark comes from a streamer breakdown in the space-charge distorted gap. Thus while in plane parallel geometry the threshold for breakdown is set by that for a Townsend discharge, the mechanism at all times is a streamer mechanism. In overvolted plane parallel gaps and in all asymmetrical gap breakdown the breakdown begins with streamer formation, and in some cases the sparking threshold is set by the threshold for a streamer formation.

While both anode and cathode directed streamers occur in midgap breakdown with heavily overvolted plane parallel gaps, by all odds the most effective streamer process, in virtue of its conservative nature, is the cathode-directed positive streamer. It is a study of the more detailed properties of this streamer process that forms the basis of this paper.

Such streamers have been studied in the positive point stepped leader strokes of lightning discharges by McEachron,[7] in the positive point-to-plane impulse breakdown streamer by J. M. Meek and R. Saxe, and recently for the preonset positive corona streamers by Amin[8] in the writer's laboratory. These studies yield certain quantitative data from which the inferences in this article derive.

In brief the mechanism of the streamer process is as follows: Assume that near the anode the geometrically determined field distribution, and the density and mean free path of gas ionizing photons created by the avalanche are adequate. Then the arrival of one adequate electron avalanche, or a suitable convergence of avalanches at the anode, leading to a limiting accumulation of positive ion space charge density and a sufficient number of active photons, permit a streamer to start. Under these conditions the vector addition of the space charge field and the initiating anode gradient are such as to cause the photoelectrons generated by the avalanches to initiate new avalanches along the field axis as these converge on the space charge. Thus the space charge will propagate itself outward towards the cathode as a positive streamer. The important factors in the advance are adequate photon density, the strength of the space charge tip field, and its extent relative to the photoionizing free path. The idealized quantitative threshold for such a process has been derived by the writer and Wijsman.[9]

If the streamer tip field is maintained adequate by current supply from the anode and by appropriate field conditions in the gap, it will advance across the gap to

* Some of the investigations underlying this paper were supported by the U. S. Office of Naval Research and by the Research Corporation of America.

[1] E. O. Lawrence and F. G. Dunnington, Phys. Rev. **35**, 396 (1930).

[2] H. Raether, Ergeb. exakt. Naturwiss. **22**, 86 (1949); H. Raether and E. Flegler, Z. tech. Phys. **16**, 436 (1935); Z. Physik **103**, 315 (1936); **104**, 219 (1936). H. Raether, Z. Physik **112**, 463 (1939).

[3] T. E. Allibone and J. M. Meek. Proc. Roy. Soc. (London) **A166**, 97 (1938); **A169**, 246 (1938).

[4] J. M. Meek and R. F. Saxe, Nature **162**, 263 (1948); Allied British Industries and Research Association Report, Section *L*–Dielectrics in General, Reference *L/T* 183 (unpublished).

[5] L. H. Fisher and B. Bederson, Phys. Rev. **81**, 109 (1951); G. A. Kachickas and L. H. Fisher, Phys. Rev. **82**, 318, 519 (1951); **88**, 878 (1952); **91**, 775 (1953).

[6] H. W. Bandel, Phys. Rev. **93**, 649(A) (1954).

[7] K. B. McEachron, J. Franklin Inst. **227**, 149 (1939); also, personal communications between B. F. J. Schonland and L. B. Loeb, February 21 to August 11, 1952.

[8] M. R. Amin, J. Appl. Phys. **25**, (1954).

[9] L. B. Loeb and R. J. Wijsman, J. Appl. Phys. **19**, 797 (1948).

the anode with its relatively dense plasma of electrons and positive ions. At the point of juncture of the streamer with the cathode, the steep potential gradients resulting cause a wave of ionization to move from cathode to anode along the streamer plasma channel. This wave, called the return stroke in lightning and observed by Meek and Saxe in long sparks and in this laboratory on short ones, progresses with velocities up to 10^{10} cm/sec, the velocity depending on the potential gradient at juncture and electron density in the plasma. It increases the existing ion densities of the order of 10^{12} electrons/cm^3 in the streamer channel to the 10^{17} or more corresponding to total ionization characterizing the brilliant filamentary spark.

In what follows, it is our intention to discuss the conditions in the streamer during the advance from anode to cathode. In Fig. 1 is shown the potential fall along the streamer channel in a gap with a streamer that has nearly crossed the gap. The dashed curve A is a pictoralization of the concept which has generally been assumed under varying conditions. Here the plasma is considered very highly conducting, and calculations for example have been made as to the fields ahead of lightning stroke channels assuming it to be equipotential with the anode. Actually the measurements of Meek and Saxe[4] indicate a considerable current flow as streamers cross a 50-cm gap, so that for reasons to be given the actual conditions in the channel must be more like those shown in the solid curve B of Fig. 1. A further peculiar paradox appears in the observations of stepped leader strokes in lightning discharge, and in photomultiplier cell studies of streamers by Meek and Saxe. Here, while the actual observations of the diameter of return stroke channels in lightning show that it is small (20 cm and 0.3 cm, respectively), the diameters of the stepped leaders are some 10 m and photomultipliers show the streamers to have 20-cm diameter. Amin, on preonset streamers, observed shoulders of transient luminosity extending in time well beyond that of the sharply defined tip luminosity. While the advancing avalanche in Raether's cloud track pictures has a certain small diameter, once the anode and cathode directed streamers start to advance in midgap breakdown of overvolted gaps, there is a sudden swelling of the condensed cloud to at least twice the original avalanche radius about such streamer tip, though the later actual spark channels are narrow. The interpretation of these contradictory observations constitute the purpose of this article.

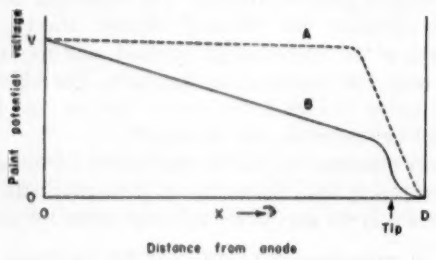

FIG. 1. Potential V of the breakdown streamer channel as a function of distance x from the anode in a gap of length D. Dashed curve A as frequently depicted in the past. Full curve B as indicated by observations of Meek and Saxe and Amin.

Experiment permits measurment of the time between initiation of the streamer at the anode (electrically indicated on the oscillograph) and its arrival at a slit normal to the streamer path and at a measured distance from the anode viewed by photomultiplier and recorded by oscillograph. The plot of distance-time curves at once yields the value of the tip velocity v_t at any point of streamer advance. Simultaneous measurement of current to the anode yields the current i at any distance x from the anode, so that current and tip velocity can be related at that point. Meek and Saxe observed that, except at the start, current and tip velocity increased in proportion for gaps up to 50 cm as the tip advanced. For gaps greater than 50 cm, Allibone and Meek had observed the velocity at first to increase and then achieve a constant value. Amin, with preonset streamers that cross only part of the gap before positive ion space charge distortion near the plane cathode deflects and stops them, observed an initial high velocity followed by a decline as the streamer approached the end of its run. Kip,[10] English,[10] and Amin[8] for preonset streamers, and Meek and Saxe for impulse breakdown streamers for positive point in gaps under 50 cm, observed that the *quantity of charge that flowed per unit length of streamer path*, n, was remarkably constant for a given streamer type. It however varies as between lightning, breakdown, and preonset streamer types. Amin also observed that in general the intensity of luminosity produced by excitation in the corona discharges parallel ionization, and that the light production as the streamer tip advanced was remarkably constant except at the very beginning and during the cessation of advance. It is then possible to state that the quantities of net positive charge n left behind in the channel per *cm* streamer advance and v_t, the tip velocity of the streamer advance, can be measured. It is also likely that in a given streamer type both (n_e+n), the total ionization per cm, and n, the net positive charge per cm left behind in the current, are nearly constant once the streamer is properly launched in its career and advances.

The range of values of observed streamer tip velocities extends roughly from 5×10^6 to in excess of 10^8 cm/sec, with most values above 10^7 cm/sec. Now the current i flowing up the channel requires that there be a potential v_e and a potential gradient X_e active down the channel to maintain it. The drift velocity v_e of the electrons in the channel is related to the field X_e at a given pressure p by the law $v_e=A(X_e/p)^{\frac{1}{2}}$ in the range

[10] A. F. Kip, Phys. Rev. **55**, 599 (1939); W. N. English, Phys. Rev. **71**, 648 (1947).

of velocities indicated, where the values of X_c/p and v_c can be taken from observations such as those of Bradbury and Nielsen,[11] or else may be calculated roughly from theory. If v_c is set equal to v_t, then even at 10^7 cm/sec the corresponding values of X_c/p are very high, such that if gas density in the channel is constant, (no heating), the values of X_c observed integrated over the length of the streamer exceed the applied gap potential V considerably. It is thus clear that the drift velocity v_c of the electrons in the channel is less than v_t, the velocity of advance of the tip, which depends on the tip field and mean free path for photoionization. Now since the measured current $i=nev_t$ it must also be related to the number of conducting electrons n_c per unit length of streamer and the drift velocity v_c of those electrons. Hence it is necessary to write that

$$i=nev_t=n_cev_c,$$

such that

$$n=n_cv_c/v_t.$$

With $nv_t=n_cv_c$, it is seen that evaluation of n and v_t yields n_cv_c but does not yield any direct knowledge of either v_c or n_c. However, there are a number of factors relating to the values of n_cv_c and X_c that will serve to help fix n_c and v_c in order of magnitude or better. If the ion production per cm of streamer is in fact constant over most of the length, then the uniformity of the streamer channel may be assumed, so that in what follows the field strength X_c causing the current flow may also be assumed constant along the channel as in Fig. 1.

As the streamer advances, the total number of positive ions and electrons created per cm length of tip advance is given by $n+n_c$. Of these, n electrons are drawn from the channel as current, yielding the charge $ne=\int_0^t idt/x$ electrons per cm of streamer that reach the anode and a linear excess positive ion charge of n per cm along the channel. The number of electrons left behind is n_c and these carry the current along the channel as the streamer advances with a velocity v_c. In consequence of the conditions outlined, the following limiting considerations apply:

(1) Values of v_c must be such, at the existing channel temperature and constant pressure p, that the reduced value of X_c/p corresponding to the potential fall down the channel X_c obeys the condition:

$$\int_0^x X_c dx=V_x,$$

where V_x is the potential drop across the channel of length x. This sets an upper limit to v_c and a lower limit to n_c.

(2) It is clear that n_c must be at least equal to n and probably should be greater than n_c, for it is unlikely that more than half the electrons will be withdrawn from a plasma initially as dense in ionization as $n+n_c$ divided by the cross-sectional area of the channel, i.e., $\sim 10^{12}$ ions/cm^3. This yields a lower limit to n_c in relation to n.

(3) X_c must be of such value that with the n_c electrons there is enough ionization by collision present to make up losses from the channel caused by (a) diffusion (generally small in $\sim 10\mu$sec), (b) recombination of electrons and molecular ions, (dissociative recombination), and (c) electron attachment to molecules, (dissociative attachment). This condition sets a minimum value to X_c and to v_c, and a maximum value to n_c. It must also be noted that, with the possible exception of the stepped lightning leader stroke, the streamer tip and a short region immediately behind it only are luminous enough to be observed and that the streamer channels under consideration much behind the tip are nonluminous. Hence much active ionization and excitation cannot occur, and this also influences the maximum value of X_c and of temperature in the channels.

TABLE I. Estimated time in microseconds for recombinative reduction in ion concentration under various assumptions.

$\frac{n_t}{n_0}$	α cm^3/ion sec	n_0	t μsec
0.1	1×10^{-6}	10^{15}	10^{-2}
		10^{13}	1
	1×10^{-8}	10^{15}	1
		10^{13}	100
0.01	1×10^{-6}	10^{15}	10^{-1}
		10^{13}	10
	1×10^{-8}	10^{15}	10
		10^{13}	1000

The decay processes of (3) will be discussed first, since some definite decisions may be derived. Of these, only dissociative recombination and dissociative attachment are very effective in air, which will be the exemplary gas treated. Considering the recombination in air, data are as yet incomplete. Biondi and Brown,[12] in microwave plasma in O_2, observed coefficients α of the order of 10^{-7} cm^3/ion sec which might be ascribed to this loss mechanism. The loss follows a time decay according to $n_t=n_0/(n_0\alpha t+1)$, where α is the coefficient of recombination, n is the initial ion density, and n_t that after t sec. On this basis Table I has been prepared, showing the time in microseconds for decay of n_0 to n_t in ratios 0.1 and 0.01 for the possible range in values of α and ion densities n_0 found in streamer channels. It is seen that an appreciable decay of n_t/n_0 cannot be expected in streamer channels under a matter of microseconds. Since the advance of all but very long streamers

[11] R. A. Nielsen and N. E. Bradbury, Phys. Rev. **49**, 338 (1936); R. A. Nielsen, Phys. Rev. **50**, 950 (1936).

[12] M. A. Biondi and S. C. Brown, Phys. Rev. **76**, 1697 (1949).

TABLE II. Delimitation of electron concentration in streamer channels for various streamer types.

Streamer type	Net charge per cm	v_i cm/sec	n_e/n assumed	v_e cm/sec	X_c/p from v_e volts/cm per mm	$T°K$ assumed	X_c volts/cm	i amp
Corona preonset positive	8×10^{9}	4×10^{7}	8	5×10^{6}	10	300	7600	0.06
Breakdown point to plane positive	5.5×10^{12}	2×10^{6}	1	2×10^{6}	3	300	2280	2
		1×10^{7}	1	1×10^{7}	25	1200	4600	10
		1×10^{7}	2	5×10^{6}	10	470	4000	10
Stepped lightning negative	5×10^{13}	1.6×10^{8}	20	8×10^{6}	20	795	5070	1200

such as in the longest sparks and lightning discharge lies under this time scale, recombination will not figure seriously as a loss factor other than in such channels. Attachment of electrons to form O_2^- ions in the relatively hot streamer channels is unlikely because of the low energy of such ion formation. On the other hand, Harrison and Geballe[13] have observed cross sections for dissociative attachment of electrons of 3 volts energy in air with O_2 molecules to form O^- ions to be about 3×10^{-20} cm^2. The average energy of electrons in air at an X/p of the order of 20 to 30 lies around 3 ev. Attachment is detected down to an $X/p=20$, where it leads to a value of η/p of 5×10^{-3} attachments per cm advance in the field at 1 mm pressure. At $X/p=50$ the value of $\eta/p=1\times10^{-2}$. At 760 mm in air $\eta=3.8$ at $X/p=20$, so that the mean free path L for attachment is 2.6 mm. As electrons attach in such encounters according to $n_x=n_0e^{-x/L}$, the value of x to reduce n_x/n_0 to 0.1 will be 5.8 mm. With a drift velocity of 1×10^7 cm/sec, this reduces the electron concentration for electron energies of 3 volts to 0.1 in 5.8×10^{-8} sec, with electrons advancing 5.8 mm in that time. This process, then, is a serious source of electron removal as long as the average electron energy remains above 3 ev. Once the current ceases to flow and X_c/p falls below 20 or less, all electrons of the swarm above 3 ev stand a good chance of attaching as the electron energy falls below 3 ev in some 10^{-7} sec. Electrons of lower energy will not attach. The effectiveness of such attachment is clearly shown in the Trichel pulse corona at 760 mm where enough electrons attach to form O^- ions and choke off the discharge in 10^{-8} sec within 0.5 mm of the point. It should be noted that once ionization ceases with much dissociative attachment, perhaps half the negative carriers or more will be O^- ions. The ion-ion recombination with positive ions having a large cross section can, at high ion densities, produce a large decrease in carriers in intervals of microseconds.

To prevent such loss of electrons, there must be a value of X_c/p in the channel sufficient to create new electrons for all those captured and lost. Harrison and Geballe show that the number of electrons produced by electron impact per cm is less than η until X/p exceeds 32.5 in air. However, the electron current observed amplifies below this, so that some ionization by collision is detectable down to an $X/p=20$. Thus, if the density of excitation and ionization in the channel is so low that ionization of excited states, i.e., ionization by successive impacts, cannot occur, the value of X_c/p should be 20 or above for streamer channels in air. In dense and highly excited channels such as possible in the photographed stepped leader channels of lightning, X_c/p may not have to be so high.

It is now of interest to consider these limitations as applied to observed streamer processes. Before doing so it must be noted that, where currents and current densities in the streamer channel become high and streamer advance lasts of an order of a microsecond, the energy input into the channel of dense current flow can lead to an appreciable heating of the channel. Since pressure is constant, say at 760 mm, the density of the gas is reduced. This decreases the value of X_c needed to yield the effective value of X_c/p causing a drift velocity v_e. Calculations made on the basis of Meek and Saxe's data indicate that heating can take place at 40-cm advance out of the 50-cm gap length to the extent of some 1200°C. Much higher temperatures are precluded since there is no visible luminosity of the channels. The rise in temperature must come as a result of ion movement in the field through collisions

TABLE II. Radial expansion limits of various streamer channel types.

Streamer	ions/cm	Channel R radius cm	X_r kv/cm	X_r/p	Channel should expand to R' cm	Radius of observed glow in cm
Pre-onset (+)	5×10^{9}	1×10^{-2}	130	200	1.0×10^{-1}	Some expansion observed by Amin
Breakdown (+)	4.5×10^{12}	0.5	2600	3400	85	10
Stepped leader (−) or (+)	5×10^{13}	10	1500	2000	1000	100–900

[13] M. A. Harrison and R. Geballe, Phys. Rev. **91**, 1 (1953).

with molecules, only some few percent coming from excitations and ionization. The rise to 1200° would reduce the value of X_e needed at 760 mm by a factor of four. Such heating does occur in lightning streamer channels and long sparks where time intervals exceed a microsecond. In the short time intervals of Amin's preonset streamer, advance heating is unlikely.

Table II gives the results of an attempted analysis on three known types of streamers and clearly indicates the limitations considered. Unfortunately data are meagre. It is likely that the velocities of Meek and Saxe's streamers are too low as given. These were estimated from uncorrected time-distance curves of Meek and Saxe by the writer. The correction would certainly have increased the initial velocities by as much as a factor of two, and perhaps more. The data from lightning represent average values and are thus not necessarily self-consistent as to relation of charge, velocity, etc., thus yielding too high a value for the current by a factor of two. It is to be noted that Amin's streamers *which extinguish after some 1.6 cm travel* have an X_e/p of the order of 10, which is inadequate to produce much ionization. The value of X_e of 7600 volts is consistent with the potential drop during the 1-cm advance before current declines and fulfills the $\int X_e dx < V$ requirement. The underestimate of Meek and Saxe's tip velocities at the start gives low values of X_e/p and of X_e. For the same streamer, after 40 cm travel across the gap with 10 amperes of current and a velocity which could easily be 1.3×10^7 cm/sec instead of 1×10^7 cm/sec, n_e/n lies above unity and is more nearly two with X_e/p above 20 and with appropriate heating yielding $X_e=4000$, a value in harmony with the applied potential. The data in the case of the stepped lightning leaders are not seriously inconsistent and appear reasonable. Actual estimates of X_e for lightning stroke channels range from 1000 to 2000 volts/cm by Schonland,[14] to 3000 by McEachron.[7] With the high electron densities in such channels, these estimates are not unreasonable. In any event, the examples show the type of analysis possible and indicate a need for more comprehensive and accurate measurements of v_i and n for different streamers.

So far the linear number density n_e and n of carriers, or better the number of charges per cm length of streamer, have been dealt with by leaving out of consideration the volume number density. The difficulty in defining the volume density arises from the uncertain values of the diameter of the streamer channels. From theory it would be expected that the diameter of the advancing streamer tip should remain relatively small, because of the required charge density for propagation as well as because of the limited photoionizing free paths and the limited electron diffusion radius of the initiating avalanche head. Further evidence that the actual current-carrying section of the streamer is confined to narrow channels, comes from the diameter of the spark channel illuminated by the return stroke. This indicates that the region in which enough electrons exist for the rapid propagation and multiplication of the return stroke is confined to a small-radius channel along which X_e maintains the channel conducting. If now the observed radius R of the spark channel is taken as that of the streamer tip, evaluation of the quantity n at once permits calculation of the charge density and radial surface field at a distance R from the axis. Such a calculation is shown in Table III for the three types of streamers. It is seen that, with the values of n given, the radial fields at R yield values of X_r/p which are very impressive.

These fields will of course not materialize, for, as soon as X_r/p reaches values in excess of perhaps 35 in air, ionization by electron avalanches converging radially toward the streamer channel of radius R will spread the positive space charge radially outward, the electrons converging into R and some contributing to the current of n_e electrons moving up the streamer channel. Assuming that spreading ceases when X_r/p falls to 20 in the expanded channel, a radius R' of each channel can be calculated. This radius R' carries most of the *excess positive charge of n per cm* at a density of roughly 5×10^{10} ions per cm³. Now it has been observed

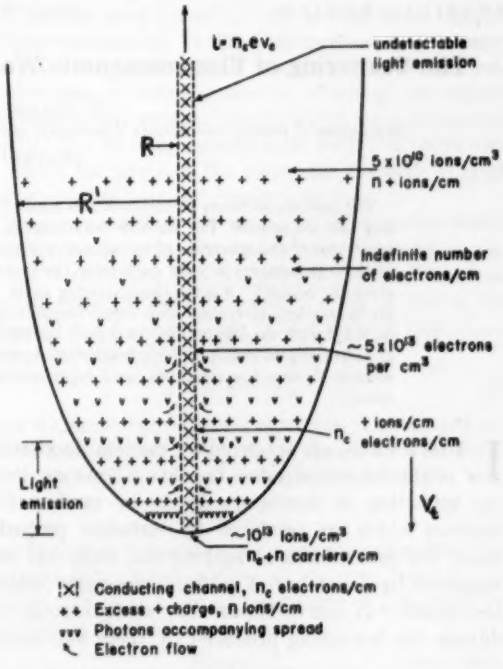

FIG. 2. Schematic drawing of the breakdown streamer tip. The central conducting channel of radius R is indicated as well as the transient luminous flaring of the excess positive space charge by radial flow of electron avalanches. The tip advances downward and the luminosity expands radially with decaying intensity from R at the tip to R' and well behind.

[14] B. F. J. Schonland (private communication).

by Meek and Saxe that, during the passage of the streamer tip in long sparks, there is a cusp-shaped luminosity that extends out radially at least 10-cm beyond the streamer axis as sensed by their photomultiplier. After this passes, the streamer channel is largely dark until the return stroke illuminates a channel of some 3-mm radius or less. In the case of lighting discharge, the radius of photographed vigorous steps with heavy currents runs out to some 10 meters. In such strokes, the channel at the cloud end of the stroke may remain visible for some little time. In this case the pilot leader invisible to the camera advances 20 to 200 meters with an active conducting channel of $R=10$ cm illuminated by the return stroke from ground some 10 *milliseconds* later, and has as well the expanded channel of positive space charge of radius R' which, with its roughly 5×10^{10} electrons per cm^3, is illuminated only by the step flash from the cloud end within some 10–100 *microseconds*. After ionization and illumination, the step ionization of radius R' decays, and it is only the original channel of $R=10$ cm which has sufficient conductivity owing to the field X_e to carry the return stroke.

It will be noted from Table III that the observed values of what might be R' are less than those computed. This is not surprising, for, as radial expansion continues, the time rate of ionization and accompanying excitation decline, and what is preceived by photomultiplier or photographic plate corresponds to values considerably less than R' depending on the sensitivity of the detector. Amin could observe only at a value of radius greater than 0.05 cm from the streamer axis. That he did not observe luminosity between 0.05 and 0.1 cm is not surprising. That, however, there was a transient luminosity after the intensely luminous peak passed the slit, is shown by the shoulder of luminosity following the tip and of such shape as only to be accounted for by ionization and excitation occurring long-after even an especially broad tip had passed the slit. The resolving power of Meek and Saxe's system was not such as to have revealed the details of the fine structure of the tip luminosity as observed by Amin, while their radial transient tip expansion could well be observed with their heavy currents. From the evidence presented, even though observed luminosity and calculated values R agree only in order of magnitude, it is believed that the nature of the transient tip shape and luminosity is accounted for. The transient space charge expansion described is illustrated schematically in Fig. 2, which is self-explanatory.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

The Scattering of Electromagnetic Waves by Turbulent Atmospheric Fluctuations*

F. VILLARS AND V. F. WEISSKOPF
Department of Physics and Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received December 7, 1953)

The statistical theory of turbulence is applied to the problem of density fluctuations in the troposphere and the ionosphere. For suitable wavelengths, for which the so-called similarity region (Kolmogoroff spectrum) of the spectrum of turbulence is relevant, a closed formula can be given for the scattering cross section. It contains as only parameter the turbulent power dissipation S, and its angular dependence is given by $(\sin\frac{1}{2}\theta)^{-11/3}$, θ being the scattering angle. The values of S required to explain ionospheric scattering are in excellent agreement with values found from investigations of meteor trails. Tropospheric data cannot be fitted with the assumptions of dry-air turbulence alone. The inference is that humidity fluctuations play an essential part in tropospheric scattering. A preliminary study of these latter fluctuations gives satisfactory results. Further investigations (and experimental data) are needed, however, to work out a quantitative theory.

LIGHT waves are scattered by random fluctuations of the refractive index. In what follows we derive the scattering of elementary waves by random fluctuations which are produced by turbulent perturbations. The general idea underlying this study has been suggested by a number of authors, especially Megaw[1] and Booker.[2] It will be shown that, under certain conditions, the scattering produced by these fluctuations can be expressed in terms of only one parameter, the turbulent energy S dissipated per cm^3 per sec. The conditions of validity of this relation are well fulfilled for the scattering of meter waves in the E layer of the ionosphere; for tropospheric scattering other parameters such as the inhomogeneity of potential temperature and specific humidity play an important part.

A. SIMPLE DERIVATION OF THE SCATTERING FORMULA

1. Scattering Cross Section

In this section we derive the expressions by simple qualitative arguments and leave the exact derivations for Sec. B.

* The research in this document was supported in part by the U. S. Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.
[1] E. C. S. Megaw, Nature **166**, 1100 (1950), and Proc. Inst. Elec. Engrs. (London) **100**, 7 (1953).
[2] H. Booker and W. E. Gordon, Proc. Inst. Radio Engrs. **38**, 401 (1950).

We calculate the scattering which a light beam with the wave vector $\mathbf{k}_0$ suffers in a region of space of volume V, which contains a medium whose dielectric constant ϵ fluctuates: $\epsilon=\epsilon_0+\Delta\epsilon$, where $\Delta\epsilon$ is a function of space. We assume that the time variations are sufficiently slow and do not affect the scattering. Then the scattered electric field amplitude ($\mathbf{E}=\mathbf{E}_s e^{i(\omega t-\mathbf{k}_1\cdot\mathbf{r})}$) at a distance R from the scattering volume ($R\gg V^{\frac{1}{3}}$) is given by

$$E_s=\frac{E_0}{4\pi R\lambda^2}\left|\int_V d\tau\Delta\epsilon(\mathbf{r})e^{i\mathbf{K}\cdot\mathbf{r}}\right|\sin\chi. \tag{1}$$

Here E_0 is the electric-field amplitude of the incident wave ($\mathbf{E}_{\text{inc}}=\mathbf{E}_0 e^{i(\omega t-\mathbf{k}_0\cdot\mathbf{r})}$), λ is the wavelength divided by 2π, χ is the angle between the direction of the incident field and the scattering, and

$$K=|\mathbf{K}|=|\mathbf{k}_0-\mathbf{k}_1|=2k\sin\tfrac{1}{2}\theta, \tag{2}$$

where

$$k=|\mathbf{k}_0|=|\mathbf{k}_1|=\lambda^{-1}.$$

We introduce the cross section $\sigma d\Omega$ for the scattering per unit volume of the scattering volume into the solid angle $d\Omega$:

$$\sigma d\Omega=(R^2/V)(E_s^2/E_0^2),$$

and get

$$\sigma d\Omega=|M|^2\sin^2\chi d\Omega/V\lambda^4(4\pi)^2, \tag{3}$$

with

$$M=\int d\tau\Delta\epsilon(\mathbf{r})e^{i\mathbf{K}\cdot\mathbf{r}}. \tag{3a}$$

For the ionosphere, we assume that $\Delta\epsilon$ is due to fluctuations in electron concentration, proportional to the density fluctuations $\Delta\rho$ of the carrier medium:

$$\Delta\epsilon=g\Delta\rho/\rho, \tag{4}$$

and

$$g=-\omega_N^2/\omega^2;\quad \omega_N^2=(1/\epsilon_0)(Ne^2/m), \tag{5}$$

where N is the density of free electrons in the medium. In the troposphere, we will have to cope with fluctuations in density, potential temperature and humidity. Because of the large dipole moment of the water molecule,[3] this latter contribution is particularly important. The dielectric constant may be written as

$$\epsilon-1=1.55\times10^{-4}\frac{p}{T}\left(\frac{\text{mb}}{^\circ\text{K}}\right)+0.74\frac{p'}{T^2}\left(\frac{\text{mb}}{^\circ\text{K}^2}\right). \tag{6}$$

In (6), p is the pressure of air, p' the partial pressure of water vapor. Since $p/T\sim\rho$, we best express $\Delta\epsilon$ in terms of density fluctuations:

$$\Delta\epsilon=0.45\Delta\rho\left(\frac{\text{g}}{\text{cm}^3}\right)+3.4\times10^3\left[\frac{\Delta\rho'}{T}-\frac{\rho'}{T^2}\Delta T\right]\left(\frac{\text{g/cm}^3}{^\circ\text{K}}\right). \tag{6a}$$

In (6a), ρ refers to the density of air, ρ' to the density of water vapor, both in g/cm³. In what follows, we shall develop a description of the density fluctuations $\Delta\rho$ from a statistical dynamics of turbulence. Since the problems involved in the discussion of the humidity fluctuations [second term in (6a)] are of rather different nature we shall discuss them separately, and presently we shall discuss the density fluctuations in dry and isentropic air only.

[3] C. M. Crain, Phys. Rev. **74**, 691 (1948).

2. Density Fluctuations in Homogeneous Turbulence

First we connect the density fluctuations with the velocity fluctuations by means of Bernoulli's law: Local velocity differences Δv will be accompanied by pressure differences Δp, and density differences $\Delta\rho$, according to

$$\Delta\rho/\rho\cong\Delta p/p=(\Delta v)^2/v_M^2, \tag{7}$$

where v_M^2 is the average square of the molecular velocity. We now concentrate our attention on the velocity fluctuations under conditions of homogeneous turbulence.[4] We may picture this situation as follows: there are external causes which constantly produce large eddies of a certain dimension L_0 and velocity v_0. These eddies divide soon into smaller eddies, say, of size $L_1=\alpha L_0$, with $\alpha<1$. The velocity of these smaller eddies embedded in the larger ones is v_1. (Measurement of v_1 is relative to the environment, i.e., relative to the motion of the large eddy.) These eddies divide again into smaller ones $L_2=\alpha^2 L_0$, with a velocity v_2 relative to their surrounding. In this dividing process, the energy fed into the largest eddy is transferred to smaller and smaller ones. A constant amount of energy per volume and time is fed into the larger eddies from an outside source. The dividing process ends with that eddy size at which the effect of the molecular viscosity is large enough to dissipate the energy into heat.

We now establish the quantitative relations which exist in this process. The energy per unit volume contained in one eddy is of the order ρv_n^2. The lifetime of such an eddy will be of the order $t_n=L_n/v_n$. Hence, the eddies of size L_n lose energy to smaller ones (size L_{n+1}) with a rate (energy/time volume):

$$S_n\sim\rho v_n^3/L_n. \tag{8}$$

Since we have a stationary process, the loss S_{n-1} of the eddies one size larger to those of size n must be equal to S_n; hence, S_n is a constant independent of n:

$$\rho v_n^3/L_n=S. \tag{8a}$$

S is the energy transmitted through the eddies from the energy source (larger eddies) all the way down to the smaller ones. The order of magnitude of S is given by the transfer from the largest eddies:

$$S\cong\rho v_0^3/L_0. \tag{8b}$$

[4] The theory of turbulence as used here was first suggested by A. N. Kolmogoroff [Compt. rend. acad. sci. U.R.S.S. **30**, 301 (1941)]. Our treatment is closest to the ideas of G. F. v. Weizsäcker [Z. Physik **124**, 614 (1948)] and W. Heisenberg [Z. Physik **124**, 628 (1948)].

We can use the magnitude of S in order to determine the size of the smallest eddies: the energy per cm³ and sec dissipated by molecular viscosity is given by $\sim\eta(dv/dx)^2$, where dv/dx is the rate of change of v per unit length. The smallest eddy is the one in which this energy reaches S. Since $(dv/dx)\sim(v_n/L_n)$, we get for the smallest eddy, whose velocity is v_S and whose L is L_S,

$$\eta v_S^2/L_S^2 \cong S.$$

Using the relation (8a): $\rho v_S^3/L_S = S$, and also (8b), we get

$$L_0/L_S = (\rho v_0 L_0/\eta)^{3/4}. \tag{9}$$

In words, the ratio of the sizes of the largest to the smallest eddies is the $\frac{3}{4}$-power of the Reynolds number, associated with the eddy-producing large scale kinematics.

We can identify the velocities v_n with the magnitude Δv used in (6). We consider regions in space of the linear size L, L fulfilling the relation

$$L_S < L < L_0. \tag{10}$$

Let us then assume that $L_n = L$; then $\Delta v = v_n$ is the average deviation from its surrounding of the velocity in that region. We then get from (6) the average density deviation $(\Delta\rho)_L$ from its surrounding in a region L:

$$(\Delta\rho)_L = \rho(v_n/v_M)^2 = \rho(v_0/v_M)^2(L/L_0)^{\frac{2}{3}}. \tag{11}$$

Let us illustrate these results by a few examples:

(a) In the *troposphere*, at 10-km height, we have $\rho = 3.6\times10^{-4}$ g cm^{-3}, $\eta = 2\times10^{-4}$ g cm^{-1} sec^{-1}. We estimate the order of magnitude of size and velocities in turbulent gusts to 1 km and 10 m/sec, respectively. Hence we may put

$$v_0 = 10^3 \text{ cm/sec}, \quad L_0 = 10^5 \text{ cm}, \tag{12}$$

and, therefore,

$$R = 1.8\times10^8, \quad L_0/L_S = 1.55\times10^6, \quad L_S = 0.65 \text{ mm},$$
$$S = 0.36 \text{ erg cm}^{-3} \text{ sec}^{-1}.$$

The smallest eddies are a fraction of a cm large.

(b) In the E-layer of the *ionosphere* at 100-km height, we have $\rho = 2\times10^{-9}$ g cm^{-3}. The viscosity is independent of pressure, and, therefore, again $\eta = 2\times10^{-4}$ g cm^{-1} sec^{-1}. From the very scant data about velocities and eddy sizes we infer roughly:

$$v_0 = 5\times10^3 \text{ cm sec}^{-1}, \quad L_0 = 5\times10^5 \text{ cm}, \tag{13a}$$

and hence[5]

$$R = 2.5\times10^4, \quad L_0/L_S = 2\times10^3, \quad L_S = 2.5\times10^2 \text{ cm},$$
$$S = 0.5\times10^{-3} \text{ erg cm}^{-3} \text{ sec}^{-1}. \tag{13b}$$

We now proceed to discuss the *humidity fluctuations*. A rough estimate of the order of magnitudes indicates that the effects of humidity in the troposphere are probably much stronger than the effects of the fluctuations of the density of the dry air. As can be inferred from (7) and (19), the average mean square fluctuation $\langle\Delta\rho^2\rangle_M$ of the latter density is of order $\rho(v_0/v_M)^4$. With the data as given by (12), and with $v_M\sim4\times10^4$ cm/sec, we get

$$(1/\rho^2)\langle\Delta\rho^2\rangle_M \sim 0.4\times10^{-6}. \tag{14}$$

We infer from (6a) that for $\rho'\sim10^{-2}\rho$, $T\sim300°$, a $\langle\Delta\rho'^2\rangle_M/\rho'^2$ or $\langle\Delta T^2\rangle_M/T^2$ of magnitude (14) produces a $\langle\Delta\epsilon^2\rangle_M$ twice as large as the dry-air density fluctuation of (14). Meteorological evidence indicates, however, that the fluctuations $\langle\Delta\rho'^2\rangle_M$ and $\langle\Delta T^2\rangle_M$ may be considerably larger: Measured $\langle\Delta T^2\rangle_M/T^2$ values are of order of magnitude 10^{-5};[6] unfortunately humidity measurements are scarce, but indicate[7] fluctuations that may be considerably larger. Practically nothing is known about the process of turbulent mixing of inhomogeneously humid air. Thus no information about short-range correlation of humidity in turbulent air is available. With some knowledge of the dynamics of turbulence it should, however, be possible to calculate approximate correlation functions from measured large-scale fluctuations (over dimensions of the order L_0). Thus, for instance, we may assume that, in an atmosphere with a negative gradient of potential temperature, convection produces large "wet" eddies (they need not be actual clouds) embedded in comparatively dryer air. The dissipation of this "surplus humidity" into the surrounding medium is then to follow closely the dissipation of the kinetic energy of the wet eddy by the breaking-up process described above. This leads to the tentative conclusion of a spectrum of humidity fluctuations very similar to the spectrum of the fluctuations of v^2, and this again is equal to the density fluctuation spectrum. Should this indeed be the case—a close study of this point is under way—the knowledge of large-scale humidity variations $(\Delta\rho')_0$ would serve to determine the humidity fluctuations $(\Delta\rho')_L$ in a volume of order L compared to its surrounding.

In fact we would get from (11)

$$(\Delta\rho')_L/(\Delta\rho')_0 = (\Delta\rho)_L/(\Delta\rho)_0 = v_n^2/v_0^2 = (L/L_0)^{\frac{2}{3}}. \tag{14a}$$

Unfortunately very little is known about the large-scale humidity fluctuations $(\Delta\rho')_0$.

3. Calculation of the Scattering Cross Section

We now determine the magnitude M appearing in (1) and defined by (2). With (4)[8] we write

$$M = \frac{g}{\rho_0}\int d\mathbf{r}\Delta\rho(\mathbf{r})e^{i\mathbf{K}\cdot\mathbf{r}}, \tag{14b}$$

[5] This estimate is in good agreement with the values given by C. deJager [Mem. soc. roy. sci., Liège, Series 4, XXI, 223 (1952)] and obtained from an analysis of meteor trails.

[6] See Electrical Engineering Research Laboratory, University of Texas, Reports Nos. 47, 1950, and 53, 1953 (unpublished).

[7] Electrical Engineering Research Laboratory, University of Texas, Reports Nos. 54, 1951, and 6-01, 1953 (unpublished).

[8] In case of tropospheric scattering, the dry-air term of (6a) gives $g = 0.45\rho(\text{g/cm}^3)$.

where ρ_0 is the average density, and we try to determine this magnitude on the basis of our picture of homogeneous turbulence. These concepts will be applicable only if the length K^{-1} lies between L_S and L_0:

$$L_S < (2k \sin\tfrac{1}{2}\theta)^{-1} < L_0. \tag{15}$$

The integral appearing in (14) is a Fourier coefficient of the density fluctuations. We can expand the density $\rho(\mathbf{r})$ in a Fourier series within the scattering volume V (which we may assume to be a cube):

$$\rho(\mathbf{r}) = \frac{1}{V} \sum_i \rho(\mathbf{k}_i) e^{-i\mathbf{k}_i \cdot \mathbf{r}}, \tag{16}$$

with the inversion

$$\rho(\mathbf{k}_i) = \int_V d\mathbf{r} \rho(\mathbf{r}) e^{i\mathbf{k}_i \cdot \mathbf{r}}. \tag{17}$$

It is then clear that M, as given by (14), is nothing but

$$M = (g/\rho_0)\rho(\mathbf{K}). \tag{18}$$

We shall now have to establish the relation between $\rho(\mathbf{k})$ and the quantity $(\Delta\rho)_L$ introduced in Sec. (2). Now $(\Delta\rho)_L$ is built up by the fluctuations whose spatial period is of order L. Because of the random nature of these fluctuations, different wavelengths will not interfere, and we have

$$(\Delta\rho)_L{}^2 = \frac{1}{V^2} \sum_{k \sim 1/L} |\rho(\mathbf{k})|^2 \cong \frac{1}{V^2} \left(\frac{V}{2\pi^2}\right) \int_{k'}^{k''} k^2 dk |\rho(\mathbf{k})|^2.$$

The interval $k''-k'$ is of order $1/L$ itself and centered about $K=1/L$. $(V/2\pi^2)k^2dk$ is the number of Fourier components with wavelength between k and $k+dk$. This gives us roughly:

$$(\Delta\rho)_L{}^2 \cong \frac{1}{6\pi^2} \frac{1}{VL^3} |\rho(1/L)|^2. \tag{19}$$

Hence,

$$|M|^2 = 6\pi^2 \left(\frac{g}{\rho_0}\right)^2 VL^3 (\Delta\rho)_L{}^2 \Big|_{L=1/K},$$

and inserting $(\Delta\rho)_L{}^2$ from (11), we get

$$|M|^2 = 6\pi^2 g^2 V \left(\frac{v_0}{v_M}\right)^4 \frac{L^{13/3}}{L_0{}^{4/3}} \Big|_{L=1/K}.$$

Inserting this into (3) gives

$$\sigma d\Omega \cong g^2 \left(\frac{v_0}{v_M}\right)^4 \frac{L^{13/3}}{L_0{}^{4/3}} \frac{\sin^2\chi}{\lambda^4} d\Omega, \tag{20}$$

where a factor of order unity (3/8) is omitted. Setting $L=K^{-1}$ and using (2) gives, finally,

$$\sigma = \frac{g^2}{L_0} \left(\frac{v_0}{v_M}\right)^4 \left(\frac{\lambda}{L_0}\right)^{1/3} (2 \sin\tfrac{1}{2}\theta)^{-13/3} \sin^2\chi. \tag{21}$$

The validity of this expression is limited by the condition (10). It is only applicable as long as

$$L_0^{-1} < 2k \sin\tfrac{1}{2}\theta < L_S^{-1}. \tag{22}$$

We find that the turbulent conditions enter the cross section only in the form $v_0{}^4/L_0{}^{4/3}$; hence, the cross section is proportional to $(S)^{4/3}$ and depends only upon the energy dissipation S.

We note the characteristic angular dependence $(\sin\frac{1}{2}\theta)^{-13/3}$. The dependence upon the wavelength is different in the troposphere and in the ionosphere. In the former, g^2 is independent of λ, and we get $\sigma \sim \lambda^{1/3}$. In the latter, we get from (5b) a wavelength dependence $\sim\lambda^{13/3}$.

If the conditions (22) are no longer fulfilled, we expect deviations from (21). The case where $L=(2k \sin\frac{1}{2}\theta)^{-1}$ is comparable to or smaller than L_S is of special interest. In this case, the eddies of that size are attenuated by molecular dissipation, and $(\Delta\rho)_L$ is then smaller than (11). In particular, it will fall off stronger than $L^{2/3}$ with decreasing L. Let us put instead of (11), for example,

$$(\Delta\rho)_L = \rho(v_0/v_M)^2 (L/L_0)^{2/3} F(L/L_S), \tag{23}$$

where $F(x)$ is a function which is equal to unity for $x \gg 1$ and which falls off rapidly for $x < 1$. We then get a factor $|F(\lambda/2L_S \sin\frac{1}{2}\theta)|^2$ multiplied into the cross section (21). Hence, the scattering falls off more strongly with decreasing λ or increasing θ if $(2k \sin\frac{1}{2}\theta)^{-1} < L_S$.

B. QUANTITATIVE THEORY

1. Cross Section

With Eqs. (1) and (2), we can write the cross section (1) as

$$\sigma d\Omega = \frac{g^2}{(4\pi)^2} \frac{1}{\lambda^4 \rho^2 V} \left| \int_V d\mathbf{r} \Delta\rho(\mathbf{r}) e^{i\mathbf{K} \cdot \mathbf{r}} \right|^2 \sin^2\chi d\Omega.$$

Now

$$\frac{1}{\rho^2} \left| \int_V d\mathbf{r} \Delta\rho(\mathbf{r}) e^{i\mathbf{K} \cdot \mathbf{r}} \right|^2 = \int d\mathbf{r} e^{i\mathbf{K} \cdot \mathbf{r}} \frac{1}{\rho^2} \int d\mathbf{R} \Delta\rho(\mathbf{R}) \Delta\rho(\mathbf{R}+\mathbf{r}) = \int d\mathbf{r} e^{i\mathbf{K} \cdot \mathbf{r}} V \frac{\langle \Delta\rho^2 \rangle_{\text{Av}}}{\rho^2} C(\mathbf{r}). \tag{24}$$

Defining the Fourier-transformed $C(\mathbf{k})$ of the correlation function $C(\mathbf{r})$ by

$$C(\mathbf{r}) = \left(\frac{1}{2\pi}\right)^3 \int d\mathbf{k} e^{-i\mathbf{k} \cdot \mathbf{r}} C(\mathbf{k}), \tag{25}$$

we get

$$\sigma d\Omega = \frac{g^2}{(4\pi)^2} \frac{1}{\lambda^4} \left(\frac{\langle \Delta\rho^2 \rangle_{\text{Av}}}{\rho^2}\right) C(\mathbf{K}) \sin^2\chi d\Omega, \tag{26}$$

with $K = 2k \sin\frac{1}{2}\theta$.

2. Velocity and Density Spectrum in the Statistical Theory of Turbulence[9]

We shall use a Fourier series expansion of the velocity field as given in Eqs. (16) and (17);[10] this enables us to write the fundamental Navier-Stokes equation in the form

$$\dot{\mathbf{v}}_{\mathbf{k}}+i\sum_{\mathbf{k}'}(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'})\mathbf{v}_{\mathbf{k}-\mathbf{k}'}+\frac{\eta}{\rho}k^2\mathbf{v}_{\mathbf{k}}+i\frac{\mathbf{k}}{\rho}p_{\mathbf{k}}=0. \tag{27}$$

We shall assume an incompressible medium

$$(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}})=0. \tag{27a}$$

This may sound paradoxical in view of our attempt to calculate the density fluctuation $\Delta\rho$. Actually our assumption merely implies that the divergence free part of the velocity field is dynamically much more important than the irrotational components. It also amounts to assuming that the potential energy stored in density variations is much smaller than the kinetic energy of the vortices. With this in mind, we shall now simply calculate the pressure fluctuations Δp, and then calculate $\Delta\rho$ with the aid of

$$p(\mathbf{r})=p_0[\rho(\mathbf{r})/\rho_0]^{\gamma}, \tag{28}$$

ρ_0, p_0 being average values of density and pressure, and $\gamma=C_p/C_v=1.4$ (for diatomic molecules). Multiplying (27) with $\mathbf{k}$ and realizing that for $k\neq 0$

$$p_{\mathbf{k}}=\Delta p_{\mathbf{k}},$$

we have

$$\Delta p_{\mathbf{k}}=-(\rho_0/k^2)\textstyle\sum_{\mathbf{k}'}(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'})(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}-\mathbf{k}'}),$$

and hence

$$\Delta\rho_k=-(\rho_0/\gamma p_0)^2k^{-2}\textstyle\sum_{\mathbf{k}'}(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'})(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}-\mathbf{k}'}). \tag{29}$$

So we now turn our attention to the velocity field $\mathbf{v}(\mathbf{k})$. Equation (27) gives, upon multiplication with $\mathbf{v}(-\mathbf{k})$,

$$-\frac{\rho}{2}\frac{d}{dt}(\mathbf{v}_{\mathbf{k}}\cdot\mathbf{v}_{-\mathbf{k}})=i\rho\textstyle\sum_{\mathbf{k}'}(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'})(\mathbf{v}_{-\mathbf{k}}\cdot\mathbf{v}_{\mathbf{k}-\mathbf{k}'})+\eta k^2(\mathbf{v}_{\mathbf{k}}\cdot\mathbf{v}_{-\mathbf{k}}). \tag{30}$$

We recognize $\frac{1}{2}\rho\mathbf{v}(\mathbf{k})\cdot\mathbf{v}(-\mathbf{k})$ as the mean kinetic energy density carried by the wave number $\mathbf{k}$, since

$$\tfrac{1}{2}\rho\langle\mathbf{v}^2(\mathbf{x})\rangle_{k_0}=\tfrac{1}{2}\rho\textstyle\sum_k(\mathbf{v}_{\mathbf{k}}\cdot\mathbf{v}_{-\mathbf{k}}).$$

Summing over all k's in (30), we get the total *power balance*:

$$S=-\frac{d}{dt}\left(\frac{\rho}{2}\textstyle\sum_k(\mathbf{v}_{\mathbf{k}}\cdot\mathbf{v}_{-\mathbf{k}})\right)$$

$$=i\rho\sum_{\mathbf{k},\mathbf{k}'}(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'})(\mathbf{v}_{-\mathbf{k}}\cdot\mathbf{v}_{\mathbf{k}-\mathbf{k}'})+\eta\textstyle\sum_{\mathbf{k}}k^2(\mathbf{v}_{\mathbf{k}}\cdot\mathbf{v}_{-\mathbf{k}}). \tag{30a}$$

The first term on the right-hand side describes the *transfer* of kinetic energy to different wave numbers $\mathbf{k}'$, the second term the *viscous dissipation* of energy. We shall now define a stationary situation within a region (L_0^3) completely enclosed in the domain over which (27) holds: this region then contains a largest eddy of linear dimension L_0 and characteristic velocity v_0. This eddy is coupled to the *outside* of that region by virtue of the "transfer" term in (30a). We assume that the effect of this coupling is to maintain the average velocity v_0 of this eddy by means of a power supply S_0. Provided S_0 is constant in time, an equilibrium will be reached in which the average $\mathbf{v}(\mathbf{k})$ is also independent of time.

From (30a) we see that S_0 must be of the form

$$S_0=\alpha\rho(v_0^3/L_0),$$

with α an absolute constant (independent of the values of v_0 and L_0), depending only on the large scale geometry of the power-supplying region outside L_0^3. Subsequently, we shall put $\alpha=1$ and *define* S_0 by

$$S_0=\rho(v_0^3/L_0). \tag{31}$$

If the Reynolds number $R_0=\rho v_0L_0/\eta$ is sufficiently large, the dissipative term will enter into play only for wave numbers $k\gg 1/L_0$. It will affect eddies of size L_S and velocity v_S, for which the "local" Reynolds number has dropped to 1:

$$\rho L_Sv_S/\eta=1. \tag{32}$$

Once wave numbers of order $1/L_S$ are reached, the frictional power dissipation replaces the power transfer to higher wave numbers. So we get another equation characterizing v_S and L_S:

$$S_0=\eta(v_S/L_S)^2. \tag{32a}$$

We shall use (32) and (32a) to *define* the two quantities v_S and L_S:

$$v_S=(S_0\eta/\rho^2)^{1/4};\quad L_S=(\eta^3/S_0\rho^2)^{1/4}. \tag{33}$$

We see that the three "external" parameters S_0, η, ρ *set an absolute scale of length and velocities* for the problem of energy transfer and dissipation in turbulent flow.

To carry this idea through in a more quantitative way, let us define a *spectral intensity* distribution $F(k)$ for the velocities:

$$\tfrac{1}{2}\textstyle\sum_{\mathbf{k}}\mathbf{v}_{\mathbf{k}}\cdot\mathbf{v}_{-\mathbf{k}}=\int_0^{\infty}dkF(k). \tag{34}$$

(In this definition, isotropy of the velocity spectrum has been assumed.) The dissipative term in (30a) is then

$$2\eta\int_0^{\infty}dkk^2F(k). \tag{35}$$

[9] Much of the content of this section is to be found in G. K. Batchelor's recent publication on the theory of homogeneous turbulence [*The Theory of Homogeneous Turbulence* (Cambridge University Press, London, 1953)] as well as Heisenberg's original paper (reference 4).

[10] We also use $v_k=v(k)$.

Since (35) represents the total power absorbed, we have, with $y=kL_S$:

$$S_0=(2\eta/L_S^3)\int_0^\infty dyy^2F(L_S^{-1}y).$$

Using (32a), we see that $F(L_S^{-1}y)=v_S^2L_SE(y)$ defines a *universal* function $E(y)$ normalized to 1:

$$1=\int_0^\infty dy2y^2E(y). \tag{36}$$

Determination of $E(y)$ or $F(k)$

(a) *Inertial range*

Provided $L_S\ll L_0$, there is a region $k\ll 1/L_S$, in which the shape of the velocity spectrum $F(k)$ is determined by the transfer mechanism and not affected by the viscous dissipation. In this range $F(k)$ should be independent of η, or $v_S^2L_SE(kL_S)$ independent of η. Clearly this can only be satisfied with a power law for E, since L_S contains η. Putting

$$E(kL_S)=\text{const}\times(kL_S)^n,$$

we get, with (33):

$$k^nv_S^2L_S^{n+1}\propto k^n\eta^{\frac{1}{2}(n+1)},$$

and independence of η gives $n=-5/3$.

To normalize F approximately, we can tentatively assume that this spectrum holds for all k values between $1/L_0$ and $1/L_S$ and that beyond both these limits the spectrum is cut off. On the high wave number side, the viscosity is certainly to act as a cutoff, whereas on the low end the power source produces naturally eddies of a certain maximum size L_0. With these assumptions the normalization condition (36) gives us:

$$\text{const}\times(2\eta/S_0)\int_{1/L_0}^{1/L_S}dkk^{\frac{1}{3}}=1.$$

(We see here that the exact shape of the spectrum at the lower end is immaterial as far as the normalization goes.) Hence,

$$\text{const}\cong\tfrac{2}{3}(S_0/\eta)L_S^{4/3}=\tfrac{2}{3}(S_0/\rho)^{2/3},$$

and

$$F(k)=\tfrac{2}{3}(S_0/\rho)^{\frac{1}{3}}k^{-5/3}. \tag{37}$$

(b) *Tail of Velocity Spectrum*

For this region no really good solution of the problem has been given. The simplest way to deal with the problem is to consider the transfer of kinetic energy from small to large wave numbers as a damping effect on the motion of the large eddies. Thus the concept of an *eddy viscosity* $\eta'(k)$ is introduced to describe the drain of energy on an eddy of size k^{-1} by eddies of size $<k^{-1}$. If we assume that $\eta'(k)$ can at all be expressed in terms of $F(k')$ $(k'<k)$, a unique expression results from dimensional arguments:

$$\eta'(k)=\text{const}\times\rho\int_k^\infty dk'[F(k')/k'^3]^{\frac{1}{2}}. \tag{38}$$

The value of the constant can be estimated from the argument that $\eta'(k)$ should be large compared with η for $k<1/L_S$ and becoming equal η at $k=1/L_S$. Hence,

$$\eta=\text{const}\times\rho\int_{1/L_S}^\infty dk'[F(k')/k'^3]^{\frac{1}{2}};$$

with (37), this gives:

$$\text{const}=(4/3)\sqrt{\tfrac{3}{2}}\cong 1.$$

We can now picture our stationary power transfer as a dissipation process whereby the energy S_k dissipated in eddies $k'\leq k$ *or* transferred to the velocity components $v_{k'}$, with $k'>k$ is given by

$$S_k=2[\eta'(k)+\eta]\int_0^k dk'k'^2F(k'). \tag{40}$$

The condition of stationary transfer implies that no energy is accumulated in a given wave number interval and hence that S_k is *independent of* k and equal to S_0. With (38) we thus get the following equation for $F(k)$:

$$S_0=2\left\{\rho\int_k^\infty dk'[F(k')/k'^3]^{\frac{1}{2}}+\eta\right\}\int_0^k dk''k''^2F(k''). \tag{41}$$

The solution of this equation is found to be[11]

$$F(k)=(8S_0/9\rho)^{2/3}k^{-5/3}(1+(k/k_S)^4)^{-4/3}, \tag{42}$$

with

$$k_S=(3S_0\rho^2/8\eta^3)^{1/4}.$$

As we see, for $k\ll k_S$, we find again our result (37) with a slight—and for our purpose, immaterial—change of the normalization constant. Also the value k_S is essentially identical with $1/L_S$ as given in (33). For $k\gg k_S$, $F(k)$ behaves like k^{-7}:

$$F(k)\cong 8^{-2/3}(S_0\rho/\eta)^2k^{-7}. \tag{43}$$

It might be interesting to compare here k_S with $k_0=1/L_0$ and $k_M=1/L_M$, L_M being the *mean free path* of the molecules of the gas: From simple gas kinetics, we have

$$(\eta/\rho)=(2\pi/3)v_M/k_M,$$

and

$$R=v_0L_0/v_ML_M \quad \text{(Reynolds number)}.$$

From (33) we then obtain:

$$k_S=R^{1/3}k_0, \tag{44a}$$

[see also (9)] and

$$k_M=(S_M/S_0)k_S, \tag{44b}$$

[11] The solution of Eq. (41) is due to J. Bass, Compt. rend. **228**, 22 (1949).

where we have introduced $S_M=\rho v_M^3 k_M$ in analogy to S_0.

Equation (44a) shows that $k_S \gg k_0$ if only the Reynolds number is large enough.

Equation (44b) decides the more subtle question whether hydrodynamic concepts apply at all in the discussion of the dissipation process. $k_M \gg k_S$ is a necessary condition for it, and it holds only provided the power supply S_0 is not too large. We shall discuss the numerical aspect of these conditions in Sec. C.

Our analysis of the velocity spectrum is now essentially completed, and we shall presently return to a discussion of the density fluctuations as given by Eq. (29).

The correlation function $C(r)$ introduced in (24) may be expressed in terms of the ρ_k ($\rho_k=\Delta\rho_k$ for $k\neq 0$):

$$\langle\Delta\rho^2\rangle_{Av} C(\mathbf{r})=\langle\Delta\rho(\mathbf{R})\Delta\rho(\mathbf{R}+\mathbf{r})\rangle_{Av}$$

$$=\frac{1}{V}\int_V d\mathbf{R}\Delta\rho(\mathbf{R})\Delta\rho(\mathbf{R}+\mathbf{r})$$

$$=\sum_{\mathbf{k}}\langle\rho_{\mathbf{k}}\rho_{-\mathbf{k}}\rangle_{Av} e^{i\mathbf{k}\cdot\mathbf{r}}.$$

[Note: The remaining averages—once the space averages are carried out—are time (phase) averages.] With (29) and $\langle\rho^2\rangle_{Av}\cong\rho_0^2$, we get

$$C(\mathbf{r})=\left(\frac{\rho_0}{\gamma p_0}\right)^2\sum_{\mathbf{k}} k^{-4}e^{i\mathbf{k}\cdot\mathbf{r}}\sum_{\mathbf{k}'\mathbf{k}''}\langle(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'})(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}-\mathbf{k}'})$$

$$\times(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}''})(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}''-\mathbf{k}})\rangle_{Av}.$$

The sum $\sum_{\mathbf{k}'\mathbf{k}''}$ can be collapsed if we notice that only the combinations

$$\text{(a)}\quad \mathbf{k}'=\mathbf{k}'',\quad \mathbf{k}-\mathbf{k}'=\mathbf{k}-\mathbf{k}'',$$

$$\text{(b)}\quad \mathbf{k}'=\mathbf{k}-\mathbf{k}'',\quad \mathbf{k}-\mathbf{k}'=\mathbf{k}''$$

give nonzero contributions in the average. Thus $\sum_{\mathbf{k}'\mathbf{k}''}$ reduces to

$$2\sum_{\mathbf{k}'}\langle(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'})(\mathbf{k}\cdot\mathbf{v}_{-\mathbf{k}'})\rangle_{Av}\times\langle(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}-\mathbf{k}'})(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'-\mathbf{k}})\rangle_{Av}.$$

Furthermore, since the incompressibility implies $(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}})=0$, we have:

$$\langle(\mathbf{k}\cdot\mathbf{v}_{\mathbf{k}'})(\mathbf{k}\cdot\mathbf{v}_{-\mathbf{k}'})\rangle_{Av}=\langle(\mathbf{k}_\perp\cdot\mathbf{v}_{\mathbf{k}'})(\mathbf{k}_\perp\cdot\mathbf{v}_{-\mathbf{k}'})\rangle_{Av}=\tfrac{1}{2}|\mathbf{v}_{\mathbf{k}'}|^2 k_\perp^2,$$

where

$$\mathbf{k}_\perp=\mathbf{k}-\mathbf{n}'(\mathbf{k}\cdot\mathbf{n}'),\quad \mathbf{n}'=\mathbf{k}'/k',$$

and, therefore,

$$k_\perp^2=[k^2k'^2-(\mathbf{k}\cdot\mathbf{k}')^2]/k'^2.$$

This gives

$$(\langle\Delta\rho^2\rangle_{Av}/\rho^2)C(\mathbf{r})$$

$$=2(\rho_0/\gamma p_0)^2\sum_{\mathbf{k}} k^{-4}e^{i\mathbf{k}\cdot\mathbf{r}}\sum_{\mathbf{k}'}\langle|\mathbf{v}_{\mathbf{k}'}|^2\rangle_{Av}\langle|\mathbf{v}_{\mathbf{k}-\mathbf{k}'}|^2\rangle_{Av}$$

$$\times\frac{[k^2k'^2-(\mathbf{k}\cdot\mathbf{k}')^2]^2}{4k^2(\mathbf{k}-\mathbf{k}')^2}.$$

Since we need only the Fourier-transformed $C(\mathbf{k})$ of $C(\mathbf{r})$, [see (25) and (26)], we write the $\mathbf{k}$ sums in terms of Fourier integrals with the aid of the substitution

$$\sum_{\mathbf{k}}\to V(2\pi)^{-3}\int d\mathbf{k}.$$

Introducing at the same time $F(k)$ by (34a), we get

$$(\langle\Delta\rho^2\rangle_{Av}/\rho^2)C(\mathbf{r})$$

$$=\pi(\rho_0/\gamma p_0)^2(2\pi)^{-3}\int d\mathbf{k}k^{-4}e^{i\mathbf{k}\cdot\mathbf{r}}$$

$$\times\int d\mathbf{k}'F(k')F(|\mathbf{k}-\mathbf{k}'|)\left(\frac{k^2k'^2-(\mathbf{k}\cdot\mathbf{k}')^2}{k^2(\mathbf{k}-\mathbf{k}')^2}\right)^2,$$

or, in view of (26):

$$\langle\Delta\rho^2\rangle_{Av}/\rho^2 C(\mathbf{k})=\pi(\rho_0/\gamma p_0)^2 k^{-4}\int d\mathbf{k}'F(k')$$

$$\times F(|\mathbf{k}-\mathbf{k}'|)\left(\frac{k^2k'^2-(\mathbf{k}\cdot\mathbf{k}')^2}{k^2(\mathbf{k}-\mathbf{k}')^2}\right)^2. \quad (45)$$

We shall first evaluate (45) for a case

$$k_0\ll k\ll k_S. \quad (46)$$

In that case, we shall use the spectrum $F(k)$ as given by (37) and use a straight cutoff at $k\geqslant k_S$, as well as for $k\leqslant k_0$. We shall see, however, that with the assumptions of (46), the k' integral is essentially independent of the two limits. The k' integral in (45)—let us call it $J(k)$—is most easily evaluated in bipolar coordinates:

$$k'=kr_1,\quad |\mathbf{k}-\mathbf{k}'|=kr_2,\quad d\mathbf{k}'=2\pi k^3 r_1 r_2 dr_1 dr_2.$$

This gives for $J(k)$:

$$J(k)=(8S_0/9\rho)^{4/3}(2\pi k^{-1/3})\int\int dr_1 dr_2 (r_1r_2)^{-14/3}$$

$$\times[(r_1^2-r_2^2)^2-2(r_1^2+r_2^2)+1]^2/16$$

$$=(8S_0/9\rho)^{4/3}(2\pi k^{-1/3})H(k_0/k;\,k_S/k). \quad (47)$$

H is a dimensionless integral and with (46) approximately an absolute constant. So we have

$$(\langle\Delta\rho^2\rangle_{Av}/\rho^2)C(\mathbf{k})\cong 2\pi^2(\rho_0/\gamma p_0)^2(S_0/\rho)^{4/3}k^{-13/3}H. \quad (48)$$

An approximate integration gives for H, assuming (46),

$$H=1.2+1.6(k/2k_S)^{7/3}+\cdots. \quad (49)$$

(In all practical cases, k is probably very much larger than k_0 and not so far from k_S. We, therefore, did not evaluate corrections $\propto k_0/k$ in H.)

In the limit of very high wave numbers k,

$$k\gg k_S, \quad (50)$$

the main contribution to $J(k)$ will arise from the shaded regions of Fig. 1. So we may write

$$J(k)=\int d\mathbf{k}'F(k')F(|\mathbf{k}-\mathbf{k}'|)\sin^4\alpha$$
$$\cong 2\int_{k'<k_S} d\mathbf{k}'F(k')F(k)\sin^4\alpha.$$

Introducing again dimensionless variables $\mathbf{z}$: $\mathbf{k}'=k\mathbf{z}$, we get

$$J(k)\cong 2(8S_0/g\rho)^{4/3}k^{-1/3}(k_S/k)^{16/3}\int_{z\leq k_S/k} d\mathbf{z}z^{-5/3}\sin^4\alpha,$$

with

$$k_S\gg k \quad (50), \quad \sin^2\alpha\cong\sin^2\varphi \text{ (see Fig.1)},$$

and hence

$$\int d\mathbf{z}z^{-5/3}\sin^4\alpha\cong 2\pi(3/4)(k_S/k)^{4/3}\times(16/15),$$

and

$$J(k)\cong(8S_0/9\rho)^{4/3}2\pi k^{-1/3}\times(8/5)(k_S/k)^{20/3}. \quad (51)$$

So $H(k_S/k)$ is $8/5(k_S/k)^{20/3}$, compared with 1.2 for the case $k\ll k_S$. We see that the two expressions (47) and (51) for $H(k_S/k)$ give almost identical values for $k=k_S$:

$$H=1.2 \text{ and } 1.6, \text{ respectively.}$$

We, therefore, conclude that the formulas hold up to (or down to) values rather close to that limit and that an interpolation should be easy.

Let us finally write down the cross-section formula corresponding to the two cases $k\ll k_S$ and $k\gg k_S$.

According to (48) and (16) we have

$$\sigma d\Omega=\frac{g^2}{8\lambda^4}(\rho_0/\gamma p_0)^2(S_0/\rho)^{4/3}K^{-13/3}H(K/k_S)\sin^2\chi d\Omega.$$

Now $(\rho_0/\gamma p_0)=(3/\gamma v_M^2)$, where v_M^2 is the mean square molecular velocity, and $\gamma=1.4$ for oxygen and nitrogen. Then finally, with $K=2k\sin\frac{1}{2}\theta$,

$$\sigma d\Omega\cong\tfrac{1}{2}g^2(S_0^{4/3}v_M^{-4})$$
$$\times\lambda^{1/3}(2\sin\tfrac{1}{2}\theta)^{-13/3}H\left(\frac{k}{2}\sin\tfrac{1}{2}\theta\right)\sin^2\chi d\Omega.$$

So, for the two cases, by using the extreme values (49) and (51) for H, and substituting S_0 from (31), we finally get

$$\sigma\cong 0.6g^2(v_0/v_M)^4\lambda^{1/3}L_0^{-4/3}[2\sin(\theta/2)]^{-13/3}\sin^2\chi,$$
$$[2k\sin(\theta/2)\ll k_S]; \quad (52)$$

$$\sigma\cong 0.8g^2(v_0/v_M)^9\lambda^7L_0^{-3}L_M^{-5}[2\sin(\theta/2)]^{-11}\sin^2\chi,$$
$$2k\sin(\theta/2)\gg k_S. \quad (53)$$

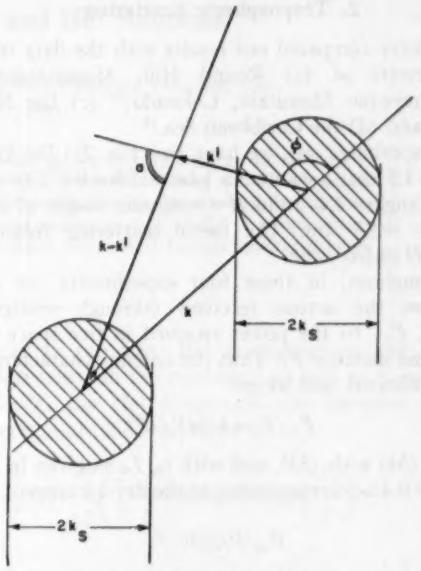

FIG. 1. The integral $J(k)$.

C. APPLICATION TO SPECIAL CASES

1. Ionospheric Scattering

We apply our scattering formula (21) to the experiment of Bailey *et al.*,[12] in which radiation of 50 Mc was scattered by the E layer. By using the same geometry as Bailey *et al.*[12] we obtain for the ratio P_r/P_t of the power received to the power transmitted:

$$\frac{P_r}{P_t}=4\sigma\frac{bA}{\sin(\theta/2)D^2},$$

where b is the thickness of the scattering layer and A is the aperture of the receiving antenna. We use the following numbers which correspond to the setup in their experiments:

$$\lambda=10^9 \text{ cm}, \quad A=3\times10^6 \text{ cm}^2, \quad D=1.2\times10^8 \text{ cm},$$
$$b=5\times10^5 \text{ cm}, \quad \theta=24°.$$

For g^2 we use (5) with $\omega_N=1.5$ Mc and the turbulence constants as given in (13a).

Substituting these values, we get

$$P_r/P_t=1.0\times10^{-18} \text{ [theoretical from (21)]},$$

whereas the experiments have given

$$P_r/P_t=0.36\times10^{-18} \text{ (experimental)}.$$

This is a remarkably good agreement. We note that $L=(2k\sin\frac{1}{2}\theta)^{-1}=250$ cm, which, according to (13b), is just equal to L_S. Hence, we actually should expect a somewhat smaller result than (21) and a steeper falling-off than (21) when going to larger angles or smaller wavelengths.

[12] D. K. Bailey *et al.*, Phys. Rev. **86**, 141 (1952).

2. Tropospheric Scattering

We have compared our results with the data of the experiments of (a) Round Hill, Massachusetts;[13] (b) Cheyenne Mountain, Colorado;[14] (c) the North Sea;[15] and (d) the Caribbean Sea.[16]

In experiment (a), we have (see Fig. 2): $D=3\times10^5$ m; $\lambda=1.5$ cm; $\alpha=$antenna lobe width$=0.6°$; $\theta=$scattering angle$=2.4°$; and $H=$minimum height of intersection $=1.8$ km. The useful scattering volume is $V\cong(\frac{1}{2}D\sin\frac{1}{2}\alpha)^2/\sin\theta$.

Throughout, in these four experiments, we shall compare the actual received (through scattering) power, P_{sc}, to the power received in free space over the same distance P_f. Thus the antenna characteristics are eliminated, and we get

$$P_{sc}/P_f=4(4\sigma V/D^2). \tag{54}$$

Using (54) with (53), and with v_0, L_0 as given in (12), and $g=0.45\rho$, corresponding to the dry-air term of (6a), we get

$$P_{sc}/P_f\cong10^{-12}.$$

The experimental data indicate a ratio of order 10^{-8}, that is, a factor 10^4 stronger.

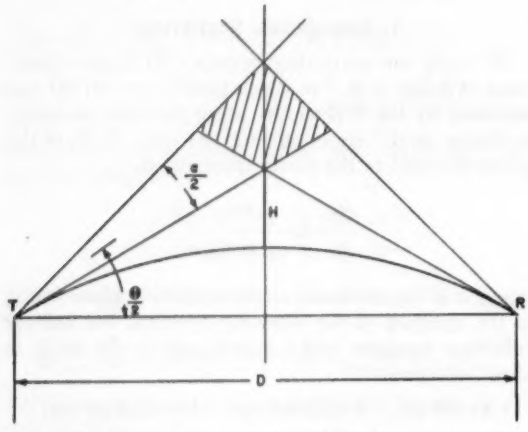

FIG. 2. Geometry for tropospheric scattering (narrow beams).

[13] Private communication.

[14] Chambers, Herbstreit, and Norton, Nat. Bur. Standards, Report No. 1826, 1952 (unpublished).

[15] E. C. S. Megaw, Nature **166**, 1100 (1950); Proc. Inst. Elec. Engrs. III, **100**, 1 (1953).

[16] For these and similar experiments, see T. J. Carroll, Nat. Bur. Standards, Report No. 1416, 1952 (unpublished).

In experiment (b) (and the following ones), we have a slightly different geometry: the beams are wide, and it is the form-factor of the scattering cross section rather than the lobe width which determines the useful scattering volume. We take, as a representative case in (*b*), a result obtained with $D=3.6\times10^5$ m, $\lambda=10$ cm, $\theta=$minimum scattering angle$=3°14'$, and $H=$minimum height of intersection $=2.54$ km, and get $P_{sc}/P_f=10^{-11}$, compared to an observed 10^{-7}. Again, we are a factor 10^{-4} off.

That the Cheyenne mountain experiments are quite typical is seen from the data from (c) and (d). The very weak λ dependence of the cross section ($\propto\lambda^{1/3}$) makes comparison easy, as far as orders of magnitudes go.

D. DISCUSSION

Further ionospheric experiments would be necessary to tell us whether the proposed picture of the dynamics of turbulence represents a reasonable approximation. Since in the troposphere the Reynolds numbers are generally much larger than in the ionosphere, we expect in general that our model is even better suited for tropospheric application. But there, in terms of dry-air fluctuations, an increase of the turbulent power output by a factor 10^8 would be necessary to account for the observed data, and this seems entirely inconceivable.

A natural explanation seems then to hold fluctuations in humidity responsible for the effect. If the assumptions made at the end of Sec. 2 are valid, we infer from (6a) and (14a) that, with say $\rho'=10^{-2}\rho$, $T\cong300°$, only a very small value for the large-scale humidity fluctuations $(\Delta\rho')_0$ is necessary to explain the observed scattering: $(\Delta\rho)_0^2/\rho'^2\cong2\times10^{-4}$. No measurements of $(\Delta\rho')_0^2$ and the humidity fluctuation correlations are available yet. Measured large-scale humidity fluctuations in and outside cumulus clouds[17] give variations in $\Delta\rho'/\rho'$ of the order of 10 percent over distances of 500 m. Fluctuations in a cloudless atmosphere are probably much smaller but might still be sufficient (1 percent) to explain the effect.

A detailed investigation of the humidity effect (and concurrently of the temperature effect) appears to be most important, and is under way.

We wish to thank Professor Henry Booker (Cornell) and Dr. J. de Bettencourt (M.I.T.) for many stimulating discussions.

[17] J. S. Malkus, Sci. American **189**, 31 (1953).

The Heat of Mixing of He^3 and He^4 Solutions

V. S. NANDA
Department of Physics and Astronomy, The Ohio State University, Columbus, Ohio
(Received December 17, 1953)

None of the current theories of He^3 and He^4 solutions is really satisfactory. In order to decide which of these makes the correct approach to the problem, experiments on the heat of mixing can be very useful. It is shown here that various theories predict remarkably different behavior for this quantity. In particular de Boer and Gorter theory gives rather high values for heat of mixing per mole of He^3 A comparison of the theoretical and experimental results should thus show clearly the need of Taconis hypothesis in explaining the various properties of He^3 and He^4 solutions.

SEVERAL preliminary attempts have been made to explain the deviation of solutions of He^3 in liquid He^4 from ideal behavior along nonclassical lines. Among these the theory of Heer and Daunt[1] gives values of vapor pressures which show the least mean square deviation from the measured values. Recently Sommers[2] has shown that reasonable agreement with the experimental vapor-pressure data is also obtained on the basis of van Laar's theory[3] of a "regular solution." Thus it is not very clear which of the two divergent opinions is the proper one, and along which direction further improvement of the theory is to be sought. Considerable information about the proper choice of the model should, however, be obtainable from measurements of the heat of mixing, since the various theories predict remarkably different behavior. Unfortunately, sufficient experimental data are not available at present to come to any decisive conclusion, but it appears worthwhile to record for future comparison some of the more important theoretical results which, incidentally, suggest some fertile regions for experimentation.

According to van Laar's theory, the solution is accompanied by a heat of mixing

$$\Delta U=\frac{N_3N_4b_3b_4}{(N_3b_3+N_4b_4)}\left(\frac{a_3^{\frac{1}{2}}}{b_3}-\frac{a_4^{\frac{1}{2}}}{b_4}\right)^2, \tag{1}$$

where N refers to the number of atoms, a and b are the usual constants occuring in van der Waal's equation of state, while the subscripts 3 and 4 refer to He^3 and He^4 components, respectively. It is obvious that ΔU is a positive quantity and is independent of temperature.

The theory of Heer and Daunt also gives a positive heat of mixing for dilute He^3 solutions. It is easily seen that for this case

$$\Delta U=\frac{3}{2}N_3kT\frac{\zeta(5/2)}{\zeta(3/2)}\frac{v_3^0}{v_4^0}\left(\frac{T}{T_0}\right)^{\frac{1}{2}}, \tag{2a}$$

where v_3^0 and v_4^0 are the atomic volumes per atom of He^3 and He^4, respectively, while T_0 is the λ temperature of the pure He^4 component. This expression holds so long as we can treat He^4 as degenerate, and He^3 as nondegenerate in the pure state as well as in solution. Above T_0 the He^4 component also becomes nondegenerate, and (2a) is replaced by

$$\Delta U=\frac{3}{2}N_3kT\frac{v_3^0}{v_4^0}\left(\frac{T_\lambda}{T}\right)^{\frac{1}{2}}\left\{0.462+0.045\left(\frac{T_\lambda}{T}\right)^1+0.0059\left(\frac{T_\lambda}{T}\right)^3+\cdots\right\}, \tag{2b}$$

where we assume $N_3\ll N_4$. Since for dilute solutions we can take $T_\lambda\approx T_0$, Eqs. (2a) and (2b) give values of ΔU over a practically continuous range of temperature.

Another theory of He^3 and He^4 solutions, due to de Boer and Gorter,[4] has also attracted a good deal of

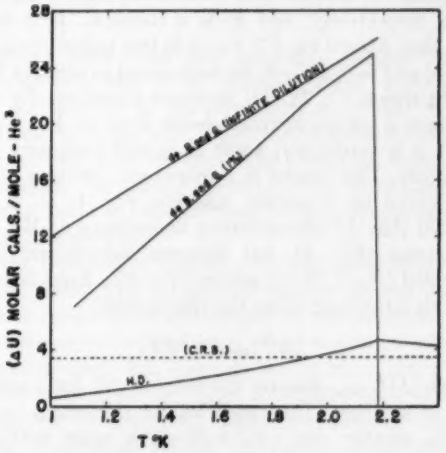

FIG. 1. The computed values of $(\Delta U)_{\text{molar}}$, the heat of mixing of liquid He^3 in liquid He^4, in dilute (He^3) solution according to (a) van Laar's theory of classical regular solutions (curve marked C.R.S.), (b) Heer and Daunt's statistical theory (curve marked H.D.), and (c) de Boer and Gorter's model (curves marked de B. and G,—infinite dilution and de B. and G., 1 percent solution).
To compute the C.R.S. curve, Eq. (1) was used with $a_4=2.77\times10^4$, $a_3=2.21\times10^4$ (cm³/mole)², $b_4=19.2$, $b_3=23.8$ cm³/mole. For curve H.D., Eq. (2) was plotted with $Nv_4^0=27.6$ cc, and $Nv_3^0=37.6$ cc (N=Avogadro's Number). See text for computations of the de B. and G. curves.

[1] C. V. Heer and J. G. Daunt, Phys. Rev. **81**, 447 (1951).
[2] H. S. Sommers, Jr., Phys. Rev. **88**, 113 (1952).
[3] J. J. van Laar, Z. physik. Chem. **72**, 723 (1910).
[4] J. de Boer and C. J. Gorter, Physica **16**, 225 (1950).

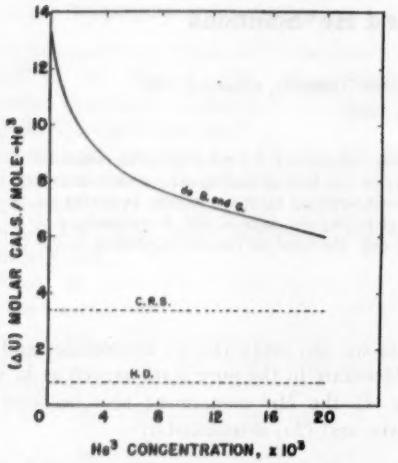

FIG. 2. The variation of $(\Delta U)_{\text{molar}}$ as a function of He^3 concentration at a constant temperature of 1.1°K according to the three models quoted in the text.

attention. Using Eqs. (8) and (32) of their paper and dropping the distinction between internal energy and enthalpy, we have in this case

$$\Delta U=\frac{N_4}{N_3+N_4}E_0(x^{7/6}-x_0^{7/6}), \tag{3a}$$

where x and x_0 denote the fraction of He^4 atoms which comprise the normal fluid in solution and in the pure state, respectively, and E_0 is a constant. It is easily seen that $\Delta U=0$ for $T\geq T_0$, as in this region $x=x_0=1$. Again, at $T=0$, $\Delta U=0$, for both x and x_0 are zero here. In the region $T\leq T_\lambda$, ΔU increases monotonically with T, since x always increases faster than x_0. For $T_\lambda<T<T_0$, x is stationary, while x_0 is still increasing continuously. This results in a monotonic decrease of ΔU (see curve for 1 percent solution, Fig. 1). It is also noticed that ΔU shows strong dependence on He^3 concentration (Fig. 2), but becomes independent of it provided $(N_3/N_4^n)\to 0$, where $N_4^n=xN_4$. Equation (3a) for this latter case takes the simple form

$$(\Delta U)_{\text{molar}}=6RT, \tag{3b}$$

where $(\Delta U)_{\text{molar}}$ denotes the value of ΔU for a mole of He^3 in solution. The condition for diluteness quoted above, namely that $(N_3/N_4^n)\to 0$, is more restrictive than the corresponding condition in the usual solution theory which only demands that $N_3/N_4\to 0$; and this is due to the necessity for the application of Taconis' hypothesis[5] on which the theory is based.

In Fig. 1 we have shown the variation of $(\Delta U)_{\text{molar}}$ with temperature for a *dilute* solution according to the three theories discussed here. The curve for 1 percent He^3 solution in de Boer and Gorter's theory is also included to show the effect of concentration on the heat of mixing in a seemingly dilute solution. In Table I the numerical values of $(\Delta U)_{\text{molar}}$ according to the three theories are given.

The following features of Fig. 1 are noteworthy: (a) The regular solution theory gives the same ΔU values even for $T>T_\lambda$. (b) Heer and Daunt's theory also leads to a finite ΔU for $T>T_\lambda$. (c) De Boer and Gorter's theory predicts rather high values of ΔU, especially near T_λ. However for $T\geq T_0$, de Boer and Gorter's model gives $\Delta U=0$. Experiment should, therefore, be able to distinguish between the applicability of all the three models. It should be noted, however, that since the Bose-Einstein model of London[6] for liquid He^4 does not give the large rate of increase of entropy between 1°K and T_0, it would be expected here also that Heer and Daunt's model, based as it is on Bose-Einstein statistics, would also underestimate the rate of rise of ΔU with temperature. It would appear, therefore, that only by appeal to experiment above T_0 as well as below T_0 could a decisive choice be made between the various models. Observations of the variation of ΔU with concentration should also lead to an unambiguous theoretical choice (see Fig. 2). So far only a single experimental observation of ΔU has been reported by Sommers[7] at 1.1°K. Exact figures, however, have not been given, though the result is claimed to be in reasonable agreement with the theoretical value obtained from (1). Without any information about the He^3 concentration (a knowledge of which is important in de Boer and Gorter's theory) and the precise measured value of ΔU, further discussion of the isolated observation cannot lead to any conclusion.*

Lastly, it may be pointed out that if we are dealing with concentrated He^3 solutions the theory of Heer and

TABLE I. Heat of mixing for a mole of He^3 as a result of forming a dilute solution in He^4.

	$(\Delta U)_{\text{molar}}$ in cal/mole He^3			
	de Boer and Gorter		Heer and Daunt	Van Laar (C.R.S.)
T°K	Dilute	1 percent	(Dilute)	(Dilute)
1.1	13.2	7.3	0.83	3.4
1.3	15.6	10.7	1.3	3.4
1.5	18.0	14.1	1.5	3.4
1.7	20.4	17.6	2.5	3.4
1.9	22.8	21.2	3.2	3.4
2.1	25.2	24.1	4.2	3.4
2.186	Zero	Zero	4.5	3.4
2.4	Zero	Zero	4.2	3.4

[5] Taconis, Beenaker, Nier, and Aldrich, Physica **15**, 733 (1949); see also J. G. Daunt, Phil. Mag. Suppl. I, 209 (1952).

[6] F. London, Phys. Rev. **54**, 947 (1938).

[7] Sommers, Keller, and Dash, Phys. Rev. **91**, 489 (1953).

* *Note added in proof.*—Sommers, Keller, and Dash [Phys. Rev. **92**, 1345 (1953)] have since published details of their experiment. They report $(\Delta U)_{\text{molar}}=1.98$ cal/mole He^3 for a 8.6 percent solution at 1.02°K. This is to be compared with the following theoretical values for $(\Delta U)_{\text{molar}}$: *CRS* = 3.02, H.D. = 0.68, de B. and G. = 3.35 cal/mole He^3. In this particular computation from Eq. (3a) we have used 'quadratic model' for Gibbs free energy of pure He^4 (see Daunt, reference 5) instead of the 'linear model' used in earlier calculations.

Daunt, unlike the other two theories, predicts a negative heat of mixing. In this case the He^3 component is degenerate in the pure state as well as in a mixture while the He^4 component, which is degenerate in the pure state, becomes nondegenerate (except for very low temperatures) in solution. For this case we have

$$\Delta U=\frac{3}{2}N_4kT\left\{1-0.514\left(\frac{T}{T_0}\right)^{\frac{1}{2}}-0.462\left(\frac{T_\lambda}{T}\right)^{\frac{1}{2}}\right.$$
$$\left.-0.0225\left(\frac{T_\lambda}{T}\right)^{3}-\cdots\right\}-\frac{2}{5}\pi\left(\frac{3}{4\pi}\right)^{\frac{1}{2}}\frac{v_4^0}{v_3^0}N_4kT^*, \quad (4)$$

where T^* is the degeneracy temperature in the pure He^3 case. From Eq. (4) we can obtain the temperature at which the heat of mixing changes sign. If the observed temperature variation of ΔU for dilute and concentrated solutions turns out to be similar in character to that predicted in Eqs. (2a), (2b), and (4), the relevance of statistics to the problem of liquid helium would be more firmly established.

It is a pleasure to record my thanks to Professor J. G. Daunt for his interest in the above investigation and for many valuable suggestions. I am also thankful to the Institute of International Education for arranging the award of a U. S. Government fellowship.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Range Measurement of Low-Voltage Electrons*†‡

MARGUERITE DAVIS
Biophysics Division, Sloane Physics Laboratory, Yale University, New Haven, Connecticut
(Received July 20, 1953; revised manuscript received December 3, 1953)

Determinations of the ranges of electrons between the energies of 600 and 2000 volts are difficult to perform since the short ranges of the particles cannot be detected by the usual physical methods. A biologically active molecule such as an enzyme can be used as detector in this energy region, the amount of inactivation in uniform layers of the enzyme being a function of the range of the electron. The ranges of the very low energy electrons determined in this manner were found to be considerably less than those predicted by the Bethe formula for energy loss if no allowance for scattering and straggling is made. If about 40 percent loss of range due to this is assumed, then theory and experiment are in fair agreement. Alternatively, a constant correction of 100A to be subtracted from the calculated range brings theory and experiment into agreement.

INTRODUCTION

THE measurement of the ranges of electrons in solids is difficult due to the statistical fluctuation of energy loss and to the multiple scattering which electrons suffer in each collision. In the low-energy region, between 600 and 2000 volts, the ranges of electrons are so small that the usual physical methods of measurement are not applicable. The use of absorbing foils is no longer feasible; the ranges are also too small to allow use of cloud chamber methods to any degree of accuracy even at 2000 volts. Since the use of low-voltage electrons as penetration probes into various kinds of structure is attractive, an experimental determination of the range of low-voltage electrons in solid material is of interest. The quantum-mechanical theory of energy loss as developed by Bethe[1] using the Born approximation assumes the interaction between the incident electron and the nucleus of the absorber atom to be weak and the probability of an electron undergoing more than one scattering process while traversing this field to be infinitely small. However, when the velocity of the particle is such that these conditions no longer hold, higher-order approximations to the solution must be used. The transition from the region where the theory of energy loss, as developed by Bethe, holds to regions where it is no longer applicable will be gradual.

In the low-energy region, the high sensitivity of biological materials to radiation can be utilized as a means of detection of the ranges of the electrons. In the present experiment, a layer of an enzyme of thickness greater than the range of the incident electrons was used as an absorber. When a uniform layer of the dry enzyme is irradiated with a sufficiently large number of electrons, all the enzyme molecules within the range of the electrons are inactivated; the enzyme beyond the range of the electrons is unaffected by the radiation. After irradiation, the material is dissolved off the cover-slip, and assayed for the activity remaining. Since the amount of enzyme originally in the sample is known, the weight of the enzyme equivalent to the fraction inactivated is easily found. The surface area of the enzyme exposed to the electron bombardment is known from the area of the cover-slip; from the

* Abstracted from a dissertation presented for the Ph.D. degree at Yale University.
† This work was performed while the author was a Predoctoral Public Health Research Fellow of the National Cancer Institute.
‡ Assisted by the U. S. Atomic Energy Commission.
[1] H. Bethe, *Handbuch der Physik* (Verlag Julius Springer, Berlin, 1933), Vol. 24, No. 1, p. 491.

amount of the material affected by the radiation, the surface area exposed, and the density of the material, taken as 1.3 g/cm³, the range of the electron is found. Since the range is calculated in the region where additional radiation of a given voltage causes no further inactivation, the range as found by this method is the distance which the electron travels in the forward direction.

EXPERIMENTAL PROCEDURE

The enzyme invertase was used in these experiments because of its stability after drying and exposure to high vacuum, and the high degree of accuracy of the assay. The assay was that of Sumner and Howell.[3] This enzyme catalyzes the hydrolysis of sugar into monosaccharides, and follows first-order reaction kinetics as long as the substrate is in excess. By providing a large excess of sugar and allowing the reaction to proceed over a long period of time, the presence of very small amounts of active enzyme can be detected with high precision. A change in activity of a few percent from a total of five micrograms of material can be measured in this manner. Under the experimental conditions used, range differences of the order of ten angstroms could be detected.

The enzyme, melabiase free, from Nutritional Biochemicals Inc., was dried down from quartz distilled water in 0.05-ml amounts onto stainless steel cover-slips, great care being taken to insure even spreading of the solution over the surface of the cover-slip. Various methods of drying were tried in order to get a uniform layer over the surface. The most consistently uniform results were found when the samples were freeze-dried. A fine crystalline network could be seen on the surface of the cover-slip. The dry samples were irradiated in vacuum (10^{-5} mm Hg pressure) by a simple electron gun consisting of a heated tungsten filament and accelerating grid. The energy of the accelerating potential measured between the filament and the irradiation plate was read on a standardized voltmeter. The amount of irradiation given to the samples in electrons/cm² was measured from the known area of electron beam and from the product of the electron current picked up on the irradiation plate times the time of irradiation. The samples dried down on the stainless steel cover-slips were placed on a metal bombardment plate in the irradiation chamber and exposed to the electrons by successively rotating the samples into the beam.

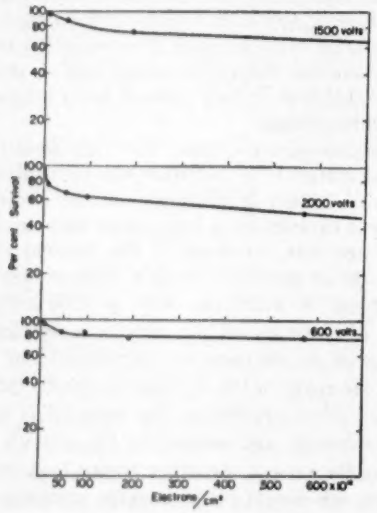

FIG. 1. Logarithm of surviving activity of invertase as a function of electrons/cm² and electron energy.

EXPERIMENTAL RESULTS

The fraction of the enzyme surviving the irradiation was investigated as a function of the number of electrons incident. Two distinct regions are apparent in the resulting curves (see Fig. 1). The initial slope is due to the inactivation of the enzyme within the range of the electrons; this trails off into a region which becomes nearly horizontal when all the molecules within the range of the electrons have been inactivated. The slight slope of this horizontal portion is presumably due to inhomogeneities in the invertase layer and to photons generated by the electron beam. The fraction which is not affected directly by the electrons can be found by extrapolating the almost horizontal part back to zero dose; this is equal to that portion of the enzyme beyond the range of the electron. From the amount inactivated, the range can be calculated by the method explained above.

Ranges were determined in this manner for electrons of energies of 100 to 2000 volts. Five micrograms of enzyme were irradiated in the 600-volt sample while 50 micrograms were used in the 1500- and 2000-volt determinations. A greater amount of scatter was found in the data at the higher energies when large amounts of enzyme were dried down on the cover-slips.

Figure 2 shows the results of the range measurements plotted against energy; the dotted line represents the ranges as calculated by Lea[3] from the Bethe theory of energy loss, assuming the validity of the extrapolation of the expression down to low energies. The values have been corrected for ranges in material of the density of protein. The constant range found experimentally from 100 to 600 volts represents the thickness of one invertase molecule. The smallest unit which can be affected corresponds to the single molecule, 90A thick. Deuteron and high-energy electron data on invertase[4] have shown the sensitive unit to be a cylinder of radius of 24A and length 83A; this corre-

[3] J. B. Sumner and S. F. Howell, J. Biol. Chem. **108**, 58 (1935).

[3] D. E. Lea, *The Action of Radiation on Living Cells* (Cambridge University Press, Cambridge, 1947), p. 23.

[4] Pollard, Powell, and Reaume, Proc. Natl. Acad. Sci. (U.S.) **38**, 173 (1952).

sponds to a molecular weight for the enzyme which is in good agreement with the value found from diffusion experiments. The present data provide further evidence that the single enzyme molecule acts as an active unit of uniform sensitivity throughout. The amount of inactivation remains constant until the incident electron has enough energy to penetrate beyond the top layer of molecules. At 600 volts the measured range of 90A can be compared with 200A, that predicted by theory. At energies greater than 600 volts, the range increases smoothly with energy. As can be seen from Table I, the difference between theory and experiment becomes less pronounced with increased energy; at 1500 and 2000 volts the discrepancy observed is a small fraction of the total range. At lower energies, the distance that the electron travels is less than the figures given by Lea. A large part of this discrepancy is probably due to scattering of the electrons. Williams[5] estimates that the penetration range is less than the path length range by about 40 percent, which would give fair agreement. The constantly decreasing discrepancy shown in the last column of Table I indicates other factors may enter, such as a constant correction the order of 100A which is to be subtracted from the calculated range. This correction could easily be due to failure of the underlying assumptions of the theory at very low electron energies.

TABLE I. Predicted and observed values of ranges of low energy electrons.

Energy of electron (volts)	Range in angstroms		Percent deviation from theory[b] (probably due to scattering)
	Theory (Bethe)	Observed[a]	
600	195	90±10	54
900	350	230±40	34
1200	545	400±80	30
1500	770	620±120	19
2000	1300	1130±225	13

[a] These values were obtained by drawing a smooth curve to fit the data shown in Fig. 2.
[b] The percentage deviation of the experimentally observed values of the range of an electron of a given energy from that predicted by theory decreases with increasing energy.

DISCUSSION

In traversing an absorber, the average electron travels a tortuous path, and it is only a rare particle that penetrates the absorber completely in the forward direction, undeviated from its original direction of motion. The range of initially monoenergetic electrons is not a clearly defined quantity as has been shown by White and Millington[6] in their measurements of the energy distribution of electrons after penetration of mica foils of various thicknesses. For a given initial energy, the number of electrons which emerge after the penetration of a foil of a given thickness steadily decreases as the thickness of the foil increases. A thickness can be found which reduces the number of electrons which penetrates to zero. This thickness is a measure of the true maximum range that is measured in the present experiment. The point where no further inactivation of the enzyme is seen for an increase in electrons incident on the enzyme can be taken as corresponding to the foil thickness that effectively stops all electrons of that energy. It has been shown by Hereford and Swann[7] that the error in range determinations in foil measurements increases with increasing energy due to the greater statistical fluctuation in energy loss per collision with electrons of higher energies.

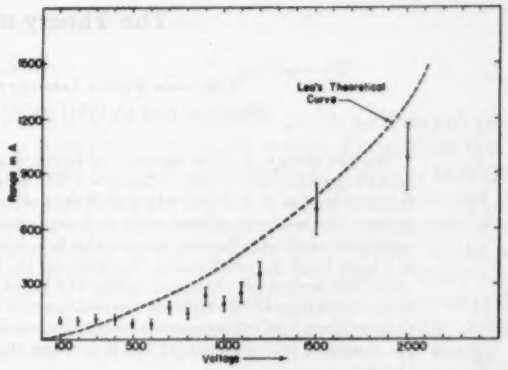

FIG. 2. Experimentally determined ranges of low energy electrons compared with theoretical values calculated by Lea from the Bethe theory of energy loss.

CONCLUSIONS

The values of ranges as determined here represent the distance which the electron travels in the forward direction. If the electron scatters in such a way that the penetration range is about 40 percent less than the path length range, the results are in fair agreement with theory. Alternatively, it may be assumed that a constant correction the order of 100A must be subtracted from the calculated ranges for electrons over 600 ev in energy.

ACKNOWLEDGMENTS

The author wishes to thank Professor E. C. Pollard for his suggestion of the problem and for his guidance and help in carrying it out, and also Dr. Franklin Hutchinson for his advice and discussions on the experiment.

[5] E. J. Williams, Proc. Roy. Soc. (London) **A130**, 310 (1930).
[6] P. White and G. Millington, Proc. Roy. Soc. (London) **A128**, 701 (1928).
[7] F. L. Hereford and C. P. Swann, Phys. Rev. **78**, 727 (1950).

The Theory of Liquid Helium

G. V. CHESTER
Wheatstone Physics Laboratory, King's College, London, England
(Received June 12, 1953; revised manuscript received August 7, 1953)

The free energy, F, of an assembly of interacting helium atoms is expressed in terms of the trace of the density operator, $\exp(-\beta H)$, where $\beta=1/kT$, H is the Hamiltonian and T the temperature. The Hamiltonian is written as H_0+gW, where H_0 is the part corresponding to the kinetic energy and gW the potential energy. The resulting expression for F is expanded in powers of g, the first two terms of the series being calculated explicitly. The first term, which is independent of g, gives the free energy of the London theory, and leads to all the usual results. In particular the transition at the lambda temperature, T_λ, is of the third order. The second term in the expansion of F raises the transition to one of the second order; all the second-order derivatives of this term are discontinuous at T_λ, while T_λ and $\partial T_\lambda/\partial\rho$ are the same as in the London theory. Numerical values are obtained for the specific heat and the discontinuity in the specific heat. These are compared with experiment and it is found that there is an improvement as compared with previous theories.

I. INTRODUCTION

THE theories of liquid helium that have so far been proposed, lead in some measure to an understanding of the lambda transition, and of the anomalous behavior below the lambda temperature. These theories are usually divided into two types according as to whether they explain the problem presented by a liquid, in terms of a gaseous model, or in terms of a solid model. They are generally associated with the names of London and Landau, respectively. Recently Temperley[1] has compared these two types of theories in a general manner. He concludes that the London type of theory is more satisfactory near the lambda temperature, while the Landau type gives better results at low temperatures, particularly near absolute zero. This is, of course, exactly what might be expected from the different assumptions inherent in the two approaches.

The original model proposed by London[2,3] was, however, very crude in that the mutual interactions between the helium atoms were not taken into account. The purpose of this paper is to present an extension of this thoery in which this defect is partly remedied. There have been several attempts to improve the original theory. The first was by London himself.[4] This was based on the assumption that the density of single-particle states, with wave vectors $\mathbf{k}$, lying between $\mathbf{k}$ and $\mathbf{k}+d\mathbf{k}$, was to depend not on k^2dk but on $k^\gamma dk$. The constant, γ, was determined so that the lambda transition became a second-order transition. The results of this assumption are too well known,[4-6] to require any elaboration. For our purpose it is sufficient to note that the assumption has never been justified theoretically. Schiff[7] showed that London's theory was the first-order term in a "quasi" virial expansion of the partition function. He calculated the next term but found that there was practically no improvement over the original theory. Such an expansion is necessarily divergent for a liquid phase in any case. Other treatments have been given by Bijl, de Boer, and Michels,[8] Goldstein,[9] and Matsubara.[10] All these theories, however, contain assumptions of an arbitrary nature, which have not so far been justified. In what follows, a new mode of analysis will be given. This consists in expanding the free energy F, in powers of a coupling constant g. The way in which g enters into the theory is readily seen if the hamiltonian of the system, H, is written in the form,

$$H=H_0+\sum_{i<j} g\omega(r_{ij}),$$
$$r_{ij}=|\mathbf{r}_i-\mathbf{r}_j|, \tag{1.1}$$

where H_0 is the kinetic energy of the system, and the other term in Eq. (1.1) is the potential energy due to the interactions between the helium atoms, the positions of which are given by the vectors $\mathbf{r}_i$. By using Eq. (1.1), the free energy may be written in the form

$$F=\sum_{n=0}^{\infty} g^n F_n, \tag{1.2}$$

where the F_n are independent of g. We shall calculate F_0 and F_1, only.

We can employ our theory to interpret the experimental facts in two different ways, which may be termed loosely the qualitative and quantitative aspects. If we make g very small then we might expect that the first two terms in Eq. (1.2) would give an accurate account of the behavior of a weakly coupled Bose-Einstein assembly. This aspect of the theory is of importance, as considerable doubt has been thrown on

[1] H. N. V. Temperley Proc. Phys. Soc. (London) **A65**, 490 (1952).
[2] F. London, Nature **141**, 643 (1938).
[3] F. London, Phys. Rev. **54**, 947 (1938).
[4] F. London, J. Phys. Chem. **43**, 49 (1939).
[5] F. London, Proceedings of the International Conference on Low Temperatures, Cambridge, 1946 (Physical Society, London, 1947).
[6] R. B. Dingle, Advances in Physics **2**, 111 (1952).
[7] L. I. Schiff, Phys. Rev. **59**, 751, 758 (1941).
[8] Bijl, de Boer, and Michels, Physica **8**, 655 (1941).
[9] L. Goldstein, J. Chem. Phys. **9**, 273 (1941); **10**, 472 (1942).
[10] T. Matsubara, Progr. Theoret. Phys. Japan **6**, 714 (1951).

London's original theory.[11,12] These doubts arise from the fact that it is difficult to see how the behavior of a liquid can be represented in any way by means of an ideal gas. Indeed, many have considered the qualitative agreement of the theory with experiment to be no more than fortuitous. Thus as our theory takes into account, in a rigorous manner, the first order departures from the ideal gas model, it should provide evidence for or against London's theory. Quantitatively we can compare the result of the theory with experimental data, using for g the value determined by de Boer[13] from calculations of the second virial coefficient of helium at higher temperatures.

In the next section the terms F_0 and gF_1 in Eq. (1.2) are calculated. The result is expressed in terms of a general interaction potential between the particles in the assembly. In carrying out this calculation in detail certain approximations are made; with these approximations it becomes obvious that the lambda temperature itself, and the dependence of the lambda temperature on the density, are the same as in London's theory.[3] In the third section the thermodynamic properties of the model are investigated from a qualitative point of view. It is found that all three second-order derivatives of F, with respect to T and V, are discontinuous at the lambda temperature. This is in agreement with experiment,[14] and marks a definite improvement over the London theory. Computations of the specific heat have been carried out for two types of interaction, and the results are compared with experiment in Sec. 4. With de Boer's potential,[13] it is found that the shape of the specific heat curve below the lambda temperature is now much nearer that found experimentally. The magnitude of the specific heat at the lambda temperature is correct, for very reasonable values of a parameter r_0 which appears in the theory. The sign of the discontinuity in the specific heat is, however, opposite to that found experimentally. If, however, we use a potential that approximates closely to the potential of the *mean* force between helium atoms, then the break in the specific heat has the correct sign, while the results below the lambda temperature are unaltered. Thus we can conclude that both qualitatively and quantitatively the model gives results which are an improvement over previous theories. Unfortunately it does not seem possible to calculate explicitly any of the higher order terms in Eq. (1.2).

II. THE EXPANSION OF THE FREE-ENERGY

Consider an assembly of N helium atoms each of mass m in a rectangular box of volume V. We write Eq. (1.1) in the form,

$$H=\sum_{i=1}^{N}\mathbf{p}_i^2/2m+gW, \tag{2.1}$$

where $\mathbf{p}_i$ is the momentum of the ith particle, and gW is the potential energy of the system. The analysis that is required, to expand F in powers of g, has been developed in a previous publication.[15] It was shown there that the partition function, and hence, the free energy, may be expanded in powers of g provided that either (a) the potential energy $\omega(r_{ij})$ has no singularities at which it becomes positively infinite, or (b) that $\omega(r_{ij})$ represents the interactions between particles that have finite incompressible cores. If, however, the potential energy is positively infinite only at isolated points then it is impossible to expand the free energy in this manner. The case where $\omega(r_{ij})$ becomes negatively infinite is of no physical interest and will not be considered. Quite generally it is clear that the potential energy of any two atoms or molecules must become very large and positive, and eventually infinite, whenever the two particles approach sufficiently close to one another. Therefore, if we wish to expand the free energy for an assembly of helium atoms in powers of g, then we must assume that they have finite incompressible cores. This seems to be perfectly justified, and such potentials have been used in many previous calculations. Thus, $\omega(r_{ij})=+\infty$ for $|\mathbf{r}_i-\mathbf{r}_j|=r_{ij}\leq 2r_0$, where r_0 is the radius of the incompressible core. We now use the formulas developed in I; the partition function Z is given by

$$Z=Z_0+gZ_1+O(g^2), \tag{2.2}$$

where, in Eq. (2.2),

$$Z_0=\sum_l G_l \exp(-\beta E_l), \tag{2.3}$$

and

$$Z_1=-\beta\sum_l G_l W_{ll} \exp(-\beta E_l). \tag{2.4}$$

In Eqs. (2.3) and (2.4), G_l is the number of states associated with the energy level E_l, and W_{ll} is given by

$$W_{ll}=\int_{S_1}\chi_l^* W \chi_l d\tau_1\cdots d\tau_N. \tag{2.5}$$

The functions χ_l in Eq. (2.5) are the correctly symmetrized solutions of the wave equation for N perfectly elastic spheres, the E_l being the corresponding eigenvalues. The above equations are equivalent to Eqs. (1.17) and (4.16) of I. The free energy F is given by

$$F=-\beta^{-1}\log Z. \tag{2.6}$$

Consequently, to the first order in g,

$$F=-\beta^{-1}\log Z_0-g\beta^{-1}Z_1/Z_0. \tag{2.7}$$

From Eqs. (2.7) and (1.2) it follows at once that

$$F_0=-\beta^{-1}\log[\textstyle\sum_l G_l \exp(-\beta E_l)], \tag{2.8}$$

[11] L. D. Landau, J. Phys. (U. S. S. R.) **5**, 71 (1941).
[12] N. N. Bogolyuboy, J. Phys. (U. S. S. R.) **11**, 23 (1947).
[13] J. H. de Boer and A. Michels, Physica **5**, 945 (1938).
[14] W. H. Keesom, *Helium* (Elsevier Publishing Company, Amsterdam, 1942).
[15] G. V. Chester, Phys. Rev. **93**, 606 (1954). This paper will be referred to hereafter as I.

and

$$F_1=\sum_l G_l W_{ll} \exp(-\beta E_l)/\sum_l G_l \exp(-\beta E_l). \quad (2.9)$$

We notice that since F_0 depends on r_0, it is *not* independent of the parameters specifying the interaction potential. It is probably impossible to evaluate either G_l or W_{ll} rigorously, and to made further progress it is necessary to find suitable approximations for them. In the case of helium the radius of the incompressible core r_0 has a value of about $1\cdot1\times10^{-8}$ cm. Furthermore the volume available to each atom V/N is, at the densities we are interested in, about 5×10^{-23} cm^3. On the other hand, the volume of the incompressible core is about 5×10^{-24} cm^3. We shall therefore make the approximation of replacing G_l in Eqs. (2.3) and (2.4) by G_l^0, its value for the free-particle system. To this degree of approximation, F_0 is the free energy of an ideal Bose-Einstein gas. We next make the approximation of replacing the elastic sphere wave functions χ_l in Eq. (2.5) by the correctly symmetrized free-particle wave functions Ψ_l. The region of integration S_1 is, for the moment, left unaltered. This insures that W_{ll} remains finite. From Eqs. (1.1) and (2.5) it is seen that W_{ll} is the sum of a number of terms of the form $\int_{S_1}\omega(r_{ij})\times d\mathbf{r}_1\cdots d\mathbf{r}_N$. In each of these terms we shall make the further approximation of extending the restricted integrations over the coordinates that do not appear in the potential function, $\omega(r_{ij})$, to cover the entire volume. The integrations over the coordinates that do appear in the potential function are still to be restricted to that region of configuration space, for two molecules, such that $r_{ij}>2r_0$. It does not seem to be worth while investigating more refined approximations until the results of these very simple ones have been established.

With these approximations, F_0 and F_1 become

$$F_0=-\beta^{-1}\log\sum_l G_l^0 \exp(-\beta E_l), \quad (2.10)$$

$$F_1=\sum_l G_l^0 W_{ll} \exp(-\beta E_l)/\sum_l G_l^0 \exp(-\beta E_l). \quad (2.11)$$

An explicit expression has been given for F_0 by London.[16] We shall therefore confine our attention to F_1. The function W_{ll}, to our degree of approximation, can easily be calculated in the usual manner. It is given by

$$W_{ll}=\tfrac{1}{2}\sum_{\mathbf{k}\neq\mathbf{k}'}\sum n_{\mathbf{k}}n_{\mathbf{k}'}K_{\mathbf{k},\mathbf{k}'}+\frac{N(N-1)}{2}K_0. \quad (2.12)$$

The symbols $n_{\mathbf{k}}$ in Eq. (2.12) are the set of numbers that define the state l. That is the state of the system is taken as being defined by the set of numbers, $(n_{\mathbf{k}})$, that specify the number of particles with wave vectors $\mathbf{k}$. The $n_{\mathbf{k}}$ run from 0 to N and satisfy the relation

$$\sum_{\mathbf{k}} n_{\mathbf{k}}=N. \quad (2.13)$$

[16] F. London, Phys. Rev. 54, 947 (1938).

The function $K_{\mathbf{k},\mathbf{k}'}$ is given by

$$K_{\mathbf{k},\mathbf{k}'}=V^{-2}\int\int_{|\mathbf{r}_1-\mathbf{r}_2|>2r_0}\omega(r_{12})\times\exp[-2\pi i(\mathbf{k}-\mathbf{k}')\cdot(\mathbf{r}_1-\mathbf{r}_2)]d\mathbf{r}_1d\mathbf{r}_2 \quad (2.14)$$

while $K_0=K_{\mathbf{k},\mathbf{k}}$ and is thus independent of $\mathbf{k}$. Combining Eqs. (2.11) and (2.12), we obtain

$$F_1=\tfrac{1}{2}\sum_{\mathbf{k}\neq\mathbf{k}'}\sum\bar{n}_{\mathbf{k}}\bar{n}_{\mathbf{k}'}K_{\mathbf{k},\mathbf{k}'}+\frac{N(N-1)}{2}K_0. \quad (2.15)$$

The $\bar{n}_{\mathbf{k}}$ are the mean values of the $n_{\mathbf{k}}$ that appear in Eq. (2.12), the means having been taken with the free particle weights $\exp(-\beta E_l)G_l^0/Z_0$. They are therefore identical with the free-particle occupation numbers of a Bose-Einstein gas, and from Eq. (2.13) satisfy the relation $N=\sum_{\mathbf{k}}\bar{n}_{\mathbf{k}}$. Now it is this relation that defines T_λ and $\partial T_\lambda/\partial\rho$, where ρ is the density. Both these quantities will consequently be identical with those of the ideal gas approximation. This last statement brings out the essential nature of the approximation of replacing G_l by G_l^0. If this had not been done, F_0 would have been dependent on r_0 and therefore the zero-order approximation to T_λ and $\partial T_\lambda/\partial\rho$ would have been dependent on r_0, and would be different from the values they have in the ideal gas approximation. Thus, if a more accurate evaluation of G_l could be carried out, it should be possible to obtain a value of T_λ nearer to that found experimentally. For convenience we list the properties of the $\bar{n}_{\mathbf{k}}$:

$$\left.\begin{aligned}&\bar{n}_{\mathbf{k}}=[\exp(\lambda k^2+\alpha)-1]^{-1},\ \text{all } \mathbf{k};\quad T>T_\lambda,\\&\bar{n}_{\mathbf{k}}=[\exp(\lambda k^2)-1]^{-1},\ \mathbf{k}>0;\quad T<T_\lambda,\\&\bar{n}_0=N-\sum_{\mathbf{k}\neq0}\bar{n}_{\mathbf{k}};\quad T<T_\lambda.\end{aligned}\right\} \quad (2.16)$$

In Eq. (2.16) $\lambda=\hbar^2/2mkT$, where k is Boltzmann's constant. The statistical parameter α is determined, as a function of T and V, for $T>T_\lambda$, by the relation $\sum_{\mathbf{k}}\bar{n}_{\mathbf{k}}=N$. For $T\leq T_\lambda$ α is identically zero. The condensation temperature T_λ is given by

$$T_\lambda=(\hbar^2/2m\pi k)[N/V\zeta(\tfrac{3}{2})]^{\frac{2}{3}}, \quad (2.17)$$

where $\zeta(\frac{3}{2})$ is the Riemann zeta function of order $\frac{3}{2}$.

It is convenient to reduce F_1 to a still more explicit form. In Eq. (2.15) we replace the $K_{\mathbf{k},\mathbf{k}'}$ by their average values, the averaging being done over all angles in $\mathbf{k}$ and $\mathbf{k}'$ space. This is legitimate because we are eventually going to replace the summations in (2.15) by integrations, and the only angular dependence comes from the $K_{\mathbf{k},\mathbf{k}'}$; the $\bar{n}_{\mathbf{k}}$ depend only on $|\mathbf{k}|$. This means that we can carry out the angular part of the integrations at once. Thus, dropping the irrelevant constant, F_1 becomes

$$F_1=\tfrac{1}{2}\sum_{|\mathbf{k}|\neq|\mathbf{k}'|}\sum\bar{n}_{\mathbf{k}}\bar{n}_{\mathbf{k}'}\langle K_{\mathbf{k},\mathbf{k}'}\rangle. \quad (2.18)$$

In Eq. (2.18),

$$\langle K_{\mathbf{k},\mathbf{k}'}\rangle=\frac{4}{V^2kk'}\int\int_{r>2r_0}\omega(r)\sin(2\pi kr)\sin(2\pi k'r)d\mathbf{r}_1d\mathbf{r}_2,$$

where $r=|\mathbf{r}_1-\mathbf{r}_2|$. The above expression can be readily reduced to the form,

$$\langle K_{\mathbf{k},\mathbf{k}'}\rangle=\frac{16\pi}{kk'V}\int_{2r_0}^{\infty}\omega(r)\sin(2\pi kr)\sin(2\pi k'r)dr. \quad (2.19)$$

The lower limit, $2r_0$, in Eq. (2.19) occurs because in Eq. (2.14) the integrations over $\mathbf{r}_1$ and $\mathbf{r}_2$ are restricted to the region $r_{12}\geqslant 2r_0$. It is now convenient to divide the calculations into two parts corresponding to temperatures below and above the lambda temperature:

(a) $T<T_\lambda$. In this range of temperature we have

$$F_1=F_1^-=\tfrac{1}{2}\sum_{|\mathbf{k}|\neq|\mathbf{k}'|}\sum\tilde{n}_{\mathbf{k}}\tilde{n}_{\mathbf{k}'}\langle K_{\mathbf{k},\mathbf{k}'}\rangle. \quad (2.20)$$

When the summations in Eq. (2.20) are replaced by integrations, care must be taken not to suppress the state $|\mathbf{k}|=0$, by using a density of states proportional to k^2, which is zero for $|\mathbf{k}|=0$. This can be avoided by writing,

$$F_1^-=\tfrac{1}{2}\sum_{|\mathbf{k}|\neq|\mathbf{k}'|\neq 0}\sum\tilde{n}_{\mathbf{k}}\tilde{n}_{\mathbf{k}'}\langle K_{\mathbf{k},\mathbf{k}'}\rangle+\tilde{n}_0\sum_{|\mathbf{k}|\neq 0}\tilde{n}_{\mathbf{k}}\langle K_{\mathbf{k},0}\rangle. \quad (2.21)$$

The summations may now be safely replaced by integrations. The states $|\mathbf{k}|=|\mathbf{k}'|\neq 0$ contribute nothing to the double integral. This is because the integrand, and all its derivatives with respect to T and V are measurable on the line $k=k'$, and these points form a set of zero (plane) measure. The quantity $\langle K_{\mathbf{k},\mathbf{k}'}\rangle$ has already been calculated in Eq. (2.19), while $\langle K_{\mathbf{k},0}\rangle$ is given by

$$\langle K_{\mathbf{k},0}\rangle=\frac{8\pi}{kV}\int_{2r_0}^{\infty}\omega(r)r\sin(2\pi kr)dr. \quad (2.22)$$

Combining Eqs. (2.22) and (2.19) with (2.21) and replacing the summations by integrations, we have

$$F_1^-=8\pi\int_{2r_0}^{\infty}\omega(r)[VI^2+\tilde{n}_0rI]dr, \quad (2.23)$$

where

$$I=\int_0^{\infty}k\sin(2\pi kr)[\exp(\lambda k^2)-1]^{-1}dk. \quad (2.24)$$

Unfortunately I cannot be calculated in terms of known functions but it can be easily expressed in the form,

$$I=(\pi/\lambda)^{\frac{1}{2}}(r/2)S_{\frac{3}{2}}^-, \quad (2.25)$$

where the functions S_μ^- have been introduced for convenience later; they are defined by,

$$S_\mu^-=\sum_{n=1}^{\infty}\exp(-\pi^2r^2/n\lambda)/n^{\mu/2}. \quad (2.26)$$

Using these results, together with the relations

$$\tilde{n}_0=N(1-\tau^{\frac{3}{2}}), \quad \text{and} \quad (\pi/\lambda_0)^{\frac{1}{2}}=N/V\zeta(\tfrac{3}{2}),$$

where

$$\lambda_0=(\lambda)_{T=T_\lambda}, \quad \text{and} \quad \tau=T/T_\lambda,$$

we have finally,

$$F_1^-=P\int_{x_0}^{\infty}x^2\omega(x)\times[(\tau^3/2\zeta)(S_2^-)^2+\tau^{\frac{3}{2}}(1-\tau^{\frac{3}{2}})S_2^-]dx, \quad (2.27)$$

where $x_0=2r_0/\sigma$, $P=4\pi RN\sigma^3/\zeta(\frac{3}{2})kV$, and R is the gas constant. The variable of integration has been changed from r to $x=r/\sigma$, and S_2^- is now given by,

$$S_\mu^-=\sum_{n=1}^{\infty}\exp(-ax^2/n)/n^{\mu/2}, \quad (2.28)$$

where $a=\pi^2\sigma^2/\lambda$.

(b) $T>T_\lambda$. We now have,

$$F_1=F_1^+=\sum_{|\mathbf{k}|\neq|\mathbf{k}'|}\sum\tilde{n}_{\mathbf{k}}\tilde{n}_{\mathbf{k}'}\langle K_{\mathbf{k},\mathbf{k}'}\rangle. \quad (2.29)$$

When we pass to the continuous-spectrum approximation, and replace the summations by integrations using a density of states proportional to k^2, we shall suppress any contributions from terms of the form $\tilde{n}_0\sum_{|\mathbf{k}|\neq 0}\tilde{n}_{\mathbf{k}}\langle K_{\mathbf{k},0}\rangle$ and $(\tilde{n}_0)^2\langle K_{0,0}\rangle$. The latter should not be present in any case, as the summation is over unequal $|\mathbf{k}|$ and $|\mathbf{k}'|$. The former should be present. It is easily seen however, that because $\tilde{n}_0$ is of order $N^{2/3}$ these terms are of order $N^{5/3}$ at most, and so are statistically negligible compared with the rest of the sum which is of order N^2. Thus it would not matter whether these terms are present or not, provided that we merely required the function F_1^+. We are, however, concerned not only with F_1^+, but also with its first and second derivatives with respect to T and V, and the derivatives of the suppressed terms are not statistically negligible at the lambda temperature; thus $\partial\tilde{n}_0/\partial T$ and $\partial^2\tilde{n}_0/\partial T^2$ are both of order N at T_λ. In Eq. (2.29) the $\tilde{n}_{\mathbf{k}}$ depend on α, Eq. (2.16), so when F_1^+ is differentiated with respect to T it will also have to be differentiated with respect to α because of the implicit dependence of α on T. Thus terms will arise in the differentiated series which were not present for $T<T_\lambda$. For example, in the first differentiation we get terms of the form $[\exp(\lambda k^2+\alpha)-1]^{-2}$. When we now pass to the continuous spectrum approximation a term such as this will give rise to an integral which diverges, as $\alpha\to 0$, in the same manner as the series $\sum_{n=1}^{\infty}e^{-n\alpha}/n^{\frac{1}{2}}$ diverges for $\alpha\to 0$. However all such integrals are multiplied by factors of the form $\partial\alpha/\partial T$,

which go to zero, for $\alpha \to 0$, just sufficiently fast for the product to yield a finite result (Appendix). It is just these terms which make up for the apparent suppression of terms like $\partial\bar{n}_0/\partial T \sum_{|\mathbf{k}|\neq 0}\bar{n}_{\mathbf{k}}\langle K_{\mathbf{k},0}\rangle$. However, terms will now arise in the double sum, which give contributions of the form $(\partial\bar{n}_0/\partial T)^2\langle K_{0,0}\rangle$ and $(\partial^2\bar{n}_0/\partial T^2)\bar{n}_0\langle K_{0,0}\rangle$. These terms should not be present since they arise from terms for which $|\mathbf{k}| = |\mathbf{k}'| = 0$. This difficulty can be easily overcome by writing Eq. (2.29) as,

$$F_1^+ = \tfrac{1}{2}\sum_{|\mathbf{k}|=|\mathbf{k}'|}\sum \bar{n}_{\mathbf{k}}\bar{n}_{\mathbf{k}'}\langle K_{\mathbf{k},\mathbf{k}'}\rangle - \tfrac{1}{2}(\bar{n}_0)^2\langle K_{0,0}\rangle, \quad (2.30)$$

where in Eq. (2.30) the summation is to be over equal values of $|\mathbf{k}|$ and $|\mathbf{k}'|$, and the second term exactly cancels out the unwanted contributions from the states $|\mathbf{k}| = |\mathbf{k}'| = 0$; the states $|\mathbf{k}| = |\mathbf{k}'| \neq 0$ give no contribution to the integral for the same reason as was given previously for the case $T < T_\lambda$. We may now safely replace the sums by integrations and get

$$F_1^+ = P\int_{x_0}^{\infty} x^2\omega(x)\tau^3/2\zeta(S_3^+)^2 dx - \tfrac{1}{2}(\bar{n}_0)^2\langle K_{0,0}\rangle, \quad (2.31)$$

where

$$S_\mu^+ = \sum_{n=1}^{\infty} \exp(-ax^2/n - n\alpha)/n^{\mu/2},$$

and τ and P have the same meaning as before.

From Eqs. (2.27) and (2.31), C_{V1}^- and C_{V1}^+ may be calculated, where C_{V1}^- and C_{V1}^+ are the contributions to C_V the specific heat at constant volume from gF_1^- and gF_1^+, respectively; as the calculations were limited to the region $T \leq T_\lambda$ we shall merely quote the result for C_{V1}^-.

$$C_{V1}^- = \frac{P\tau^{\frac{1}{2}}g}{T}\int_{x_0}^{\infty} x^2\omega(x)\left[\frac{\tau^{\frac{1}{2}}}{\zeta}f_1(x) - f_2(x)\right]dx. \quad (2.32)$$

The functions $f_1(x)$ and $f_2(x)$ are given by

$$\left.\begin{aligned} f_1(x) &= 3S_3^-(2\zeta - S_3^-) - 6zS_5^-(\zeta - S_3^-) \\ &\quad - z^2(S_5^-)^2 + z^2S_7^-(\zeta - S_3^-) \end{aligned}\right\} \quad (2.33)$$

and

$$f_2(x) = \tfrac{3}{4}S_3^- - 3zS_5^- + z^2S_7^-,$$

where

$$z = ax^2/\lambda \quad \text{and} \quad \zeta = \zeta(\tfrac{3}{2}).$$

The discontinuity in the specific heat can also be calculated from Eqs. (2.27) and (2.31). It is found to be given by

$$(\Delta C_{V1}) = (C_{V1}^+ - C_{V1}^-)_{T=T_\lambda}$$

$$= PT_\lambda^2\zeta^{-1}(\partial^2\alpha/\partial T^2)_{T=T_\lambda}\int_{x_0}^{\infty} x^2\omega(x)S_3S_{\frac{1}{2}}dx, \quad (2.34)$$

where

$$S_{\frac{1}{2}} = \sum_{n=1}^{\infty}[1 - \exp(-ax^2/n)]n^{-\frac{1}{2}},$$

$(\partial^2\alpha/\partial T^2)_{T=T_\lambda}$ is given in the Appendix, and S_3 is the common value of S_3^- and S_3^+ at T_λ, that is for $\alpha = 0$.

III. THE THERMODYNAMICS OF THE LAMBDA TRANSITION

We shall now use the expressions given by Eqs. (2.11) and (2.18) to deduce the qualitative behavior of the model at the lambda temperature. Our main interest is to determine the nature of the discontinuities in the derivatives of F. These arise from the discontinuities in the derivatives of the $\bar{n}_{\mathbf{k}}$ which in turn depend on the discontinuities in α. As the first-order derivatives of α are all continuous at the lambda temperature (Appendix), it follows that the first-order derivatives of F are continuous at this point. In particular the entropy, $S = -(\partial F/\partial T)_V$ and the pressure, $p = -(\partial F/\partial V)_T$ are continuous at T_λ. To study the second derivatives of F we require the following relations,

$$(\partial F_0/\partial\alpha) = 0, \quad (3.1)$$

and

$$(\partial F_1/\partial\alpha) \neq 0. \quad (3.2)$$

Equation (3.1) is equivalent to the relation $N = \sum_{\mathbf{k}}\bar{n}_{\mathbf{k}}$, and follows immediately when F_0 is calculated in the usual manner.[17,18] Equation (3.2) is obvious from the form of Eq. (2.18). It follows from Eq. (3.1) that the nth derivatives of F_0 depend only on the $(n-1)$th derivatives of α. Thus, as the first discontinuous derivatives of α are of the second order, the first discontinuous derivatives of F_0 will be of the *third* order; for example the gradient of the specific heat. This of course is a well known result of the ideal gas approximation. On the other hand, from Eq. (3.2), the nth derivatives of F_1 will depend on the nth derivatives of α, and hence the first discontinuous derivatives of F_1 will be of the *second* order. Now we have

$$C_V = -T\left(\frac{\partial^2 F}{\partial T^2}\right)_V; \quad \left(\frac{\partial p}{\partial T}\right)_V = -\left(\frac{\partial^2 F}{\partial T\partial V}\right);$$

$$\left(\frac{\partial p}{\partial V}\right)_T = -\left(\frac{\partial^2 F}{\partial V^2}\right). \quad (3.3)$$

Thus, the theory predicts that all three quantities on the left-hand side of relations (3.3) will be discontinuous at T_λ. This is in agreement with experiment, and represents a considerable improvement over the London theory which predicts that all these quantities should be continuous. It should be noted, however, that the experimental values correspond to discontinuities in derivatives of the *total* free energy, whereas we have only examined the behavior of the first and second terms of series (1.2) for F. But it would seem to be purely accidental if the discontinuities arising from the higher order terms in series (1.2) exactly cancelled those of the first two terms.

Our general qualitative conclusions concerning the discontinuities depend directly on the statistics of the

[17] B. Kahn, dissertation, Utrecht, 1938.
[18] A. R. Fraser, Phil. Mag. 42, 156, 165 (1951).

particles. They therefore lend considerable support to London's hypothesis that the anomalous behavior of liquid helium is closely connected with the type of statistics obeyed by the helium atoms.

IV. COMPUTATION OF SPECIFIC HEAT—COMPARISON WITH EXPERIMENT

The contribution C_{V1}^{-}, to the specific heat C_V, has been calculated for two types of interaction potential. The first type of potential was identical with that used by de Boer[18] to calculate the second virial coefficient of helium at low temperatures. That is, $g\omega(r)$ was assumed to be given by

$$g\omega(r)=g\left[\left(\frac{\sigma}{r}\right)^{12}-\left(\frac{\sigma}{r}\right)^{6}\right]; \qquad (4.1)$$

g is identical with the quantity 4ϵ in de Boer's paper, and σ is the scale factor introduced in Eq. (2.27). The constants ϵ and σ were assumed to have the same values as used by de Boer, namely $\epsilon=1.41\times10^{-15}$ erg and $\sigma=2.561\times10^{-8}$ cm. This potential was substituted into Eq. (2.32), the integrand tabulated as a function of x for various values of τ, and the integration performed numerically. The results are shown in Fig. 1, curve C, while curve A represents the experimental results.[19] The value of r_0 was chosen so that the specific heat had the correct experimental value at the lambda temperature. With a value of $r_0=1.09\times10^{-8}$ cm there is remarkably close agreement, below the lambda temperature, between the theoretical and experimental results. Moreover, this value of r_0 is in good agreement with the value of 1.1×10^{-8} cm obtained from kinetic theory. However, at the lambda temperature itself the discontinuity, calculated from Eq. (2.34), has the wrong sign and is very large ($\sim5R$). The gradient of the specific heat on the high temperature side of the lambda point has also been computed and is found to be very small and negative ($\sim-1R$). Thus, the theoretical behavior at T_λ is in complete disagreement with experiment. This is perhaps disturbing in view of the excellent agreement below the lambda temperature. With the value of r_0 that we have chosen, the specific heat becomes negative for $\tau<0.3$. This clearly indicates that the approximations we have made are breaking down in this region of temperature. This is not altogether surprising considering the nature of the approximations. It may be mentioned that with a slightly smaller value of r_0 the specific heat would be positive over the whole range.

In view of the marked disagreement with experiment at the lambda temperature, it was thought worthwhile to see if any improvement could be obtained by using a different type of potential. From the numerical calculations it was obvious that merely changing the value of r_0 would not alter the qualitative results obtained with de Boer's potential. It was however clear that if the potential was assumed to have a "repulsive hump" at a distance given by $x=1.6$, then the discontinuity in the specific heat would have the correct sign and the gradient on the high-temperature side of the lambda point would be large and negative. Now a potential of this kind corresponds closely to the potential of the *mean* force between helium atoms. Furthermore, by using this type of potential we are taking into account more accurately the fluctuations in density of the number of atoms in the neighborhood of any particular atom. Thus, although the introduction of this potential is somewhat arbitrary, it is not unreasonable on physical grounds.

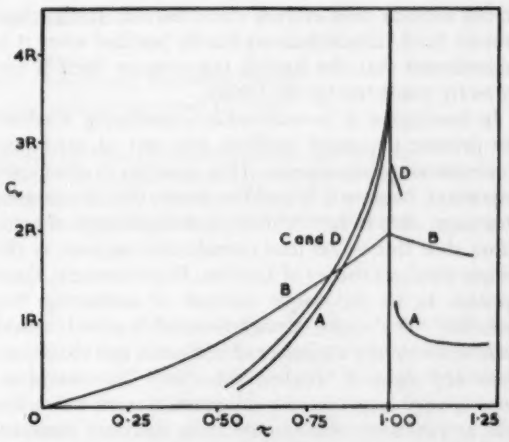

FIG. 1. The specific heat of liquid helium. A—experimental values; B—London theory; C—theory of this paper using de Boer's potential [Eq. (4.1)] for the interaction between the helium atoms; D—this paper using approximate potential of the mean force.

It was not considered worthwhile to attempt to represent this type of potential in any great analytical detail and then perform a numerical integration to obtain the specific heat. It was assumed that for our purpose a sufficiently good representation of it could be obtained by simply changing the sign in de Boer's potential at a distance given by $x=1.6$. The main defect of this potential is that it does not go to zero for large x as $-x^{-6}$, but as $+x^{-6}$. However, the contribution to the integrand is very small at these values of x and this defect will therefore not be very important. The results obtained with this potential are shown in Fig. 1, curve D. To obtain the correct value for the specific heat at T_λ, it was necessary to change r_0 slightly to a value of 1.11×10^{-8} cm. It is seen that below the lambda temperature the curve is indistinguishable from that obtained with De Boer's potential. At the lambda temperature the discontinuity in the specific heat has the correct sign, while the specific heat itself falls off rapidly above the lambda temperature. The break in the specific heat is, however, too small. It seems almost certain that with a sufficiently refined "mean force potential" the behavior

[18] W. H. Keesom and A. P. Keesom, Physica 2, 557 (1935).

at the lambda temperature could be still further improved. Such calculations are hardly justified when it is remembered that the lambda temperature itself is not correctly predicted by the theory.

In conclusion it is worthwhile considering whether the present treatment exhibits any sort of gas-liquid condensation phenomenon. This question is obviously important, because if it could be shown that the present treatment does in fact exhibit a gas-liquid type of transition then this would lend considerable support to the simple ideal gas theory of London. Unfortunately, there appears to be no simple method of answering this question. The most straightforward method would seem to be to plot a number of isotherms and see if they show any signs of "critical behavior." This would involve a very large amount of numerical work which has not, as yet, been carried out. It is, however, doubtful whether the present method would exhibit any really realistic gas-liquid transition. This is because only two terms in the expansion F have been calculated (approximately); it is very likely that many more terms in the expansion would have to be calculated before a realistic transition occurred.

The numerical work in this section was first carried out on a desk machine. It was subsequently checked by Dr. M. P. Barnett on a punched card machine, using techniques previously developed[20] for the evaluation of molecular integrals. The author wishes to thank Dr. Barnett for the care with which these calculations were carried out. He also wishes to thank Mr. F. Booth for considerable help in the preparation of the manuscript. The author is indebed to the Central Research Fund of London University for the loan of a Marchant calculating machine, and to the Department of Scientific and Industrial Research for a maintenance grant.

APPENDIX

The statistical parameter, α, is defined as a function of ρ and T (that is as a function of T/T_λ) by the equation,

$$N=\sum_k \bar{n}_k(\alpha). \tag{A.1}$$

In the continuous spectrum approximation, this becomes

$$N=(\pi/\lambda)^{\frac{1}{2}}V\sum_{n=1}^{\infty} e^{-n\alpha}n^{-\frac{1}{2}}, \tag{A.2}$$

where $\lambda=h^2/2mkT$. Equation (A.2) holds for $T>T_\lambda$; for $T=T_\lambda$, $\alpha=0$, and Eq. (A.2) becomes

$$N=(\pi/\lambda_0)^{\frac{1}{2}}V\zeta(\tfrac{3}{2}), \tag{A.3}$$

where $\zeta(\frac{3}{2})$ is the Riemann zeta function of order $\frac{3}{2}$, and $\lambda_0=(\lambda)_{T=T_\lambda}$. Dividing Eq. (A.2) by (A.3),

$$\zeta\tau^{-\frac{1}{2}}=\sum_{n=1}^{\infty} e^{-n\alpha}\cdot n^{-\frac{1}{2}}. \tag{A.4}$$

From Eq. (A.4) the derivatives of α may be calculated by direct differentiation. The calculations are made very much easier if the formula given by Robinson[21] is used. He obtains an expansion for $\sum_n e^{-n\alpha}n^{-\mu}$ of the form

$$\sum_{n=1}^{\infty} e^{-n\alpha}n^{-\mu}=\Gamma(1-\mu)\alpha^{\mu-1}+\sum_{p=0}^{\infty}\frac{(-1)^p}{p!}\alpha^p\zeta(\mu-p), \tag{A.5}$$

where $\zeta(\mu-p)$ is the Riemann zeta function of order $\mu-p$, and $\Gamma(1-\mu)$ is the gamma function of argument $1-\mu$. This expansion is valid for $|\alpha|\leq 2\pi$, and all μ. It is easily shown, using (A.5) and (A.4) that for small α $(T>T_\lambda)$,

$$\frac{\partial\alpha}{\partial T}=-\left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}}\zeta(\tfrac{3}{2})\frac{3}{2T_0}\tau^{-3/2}. \tag{A.6}$$

Thus $\partial\alpha/\partial T\to 0$ as $\alpha\to 0$. Moreover, as $\partial^n\alpha/\partial T^n \equiv 0(n\geq 1)$ for $T<T_\lambda$, it follows that $\partial\alpha/\partial T$ is continuous at $T=T_\lambda$. In a similar manner,

$$(\partial^2\alpha/\partial T^2)_{T=T_\lambda}=9[\zeta(\tfrac{3}{2})]^2/8\pi T_\lambda^2, \tag{A.7}$$

showing that $\partial^2\alpha/\partial T^2$ is discontinuous at $T=T_\lambda$. From Eq. (A.4) the following results can be easily deduced:

$$\zeta(\tfrac{3}{2})\frac{3}{2T_\lambda}=\lim_{\alpha\to 0}\sum_{n=1}^{\infty} e^{-n\alpha}n^{-\frac{1}{2}}\frac{\partial\alpha}{\partial T}, \tag{A.8}$$

and

$$\frac{3}{4}\frac{\zeta(\frac{3}{2})}{T_\lambda^2}=\lim_{\alpha\to 0}\left[\sum_{n=1}^{\infty} e^{-n\alpha}n^{\frac{1}{2}}\left(\frac{\partial\alpha}{\partial T}\right)^2-\sum_{n=1}^{\infty} e^{-n\alpha}n^{-\frac{1}{2}}\frac{\partial^2\alpha}{\partial T^2}\right]. \tag{A.9}$$

These relations have been used repeatedly in the text.

[20] M. P. Barnett and C. A. Coulson (to be published).

[21] J. E. Robinson, Phys. Rev. 83, 678 (1951).

Origin of the Ionospheric *E* Layer

K. RAWER AND E. ARGENCE
Service de Prévision Ionosphérique Militaire, Neuf-Brisach, Haut Rhin, France
(Received July 15, 1953; revised manuscript received September 11, 1953)

The intensity of solar soft x-rays (50···230A) as computed by Elwert is found to be sufficient to cause the ionization of the ionospheric *E* layer. On the other hand, preionization of molecular oxygen by ultraviolet chromospheric radiation ($\lambda<1020$A, Nicolet) could perhaps also be admitted, if it could be accepted that the dissociation of atmospheric oxygen occur in a high transition layer of considerable thickness (100–150 km). The first process is probably predominant.

1. INTRODUCTION

ACCORDING to the usual theory the *E* layer is formed by ionization of molecular oxygen by ultraviolet radiation. In 1948 Bates and Hoyle[1] proposed a theory of formation by solar x-rays at about 40A. In a first note in 1951,[2] we pointed out some difficulties of the usual theory and mentioned the x-ray theory as one possible solution. Meanwhile Elwert[3] has calculated x-ray emission from the solar corona; he found that a softer radiation (50 to 200A) should be most important for the formation of the ionospheric *E* layer.

While x-rays would cause ionization of all atmospheric molecules, the commonly accepted process should be selective and concern only oxygen molecules. Nicolet[4] suggested the wavelength range 900 to 1000A as responsible for ionization of O_2 by an assumed preionization effect.

In this situation the two theories and the corresponding deductions must be compared. This has recently been done by Choudhury[5] who calculated the absorption cross sections required in the two theories, and in more detail by the present authors.[6] We feel that a *mutual* comparison of these results together with recent developments in this field should be of interest.

2. IONIZATION OF O_2 BY ULTRAVIOLET LIGHT

If O_2 is ionized in a specific process, the shape of the corresponding layer, i.e., its height and thickness, will depend on the distribution of O_2 molecules and on the cross section of the process itself. On the other hand, the resulting ionization depends essentially on the intensity of the ionizing radiation.

Distributions of Molecular Oxygen

Many authors, including Penndorf,[7] suppose that molecules recombine by a three-body collision process. This process depends in a sensitive manner on the molecular density. The resulting transition layer is sharp (about 10 km), and its height is lower than that obtained for the *E* layer by radio measurements (maximum at about 125 km). This is an important difficulty because there is only one parameter at our disposal (i.e., the cross section of ionization) to adjust to the well-known height and thickness of the ionized layer; we can prove that this is not possible when the transition is sharp and low.[6] In our former discussion,[2] we pointed out that perhaps another recombination process should be considered. Recombination by radiative two-body collision has already been considered by Majumdar[8] who found a very high transition layer. While Rakshit[9] with the same hypothesis obtained a low and very thin transition, recently Moses and Wu[10] arrived at a layer of more than 30-km thickness situated between 100 and 130 km (see Fig. 1).

We think that the discussion on O_2 dissociation is not yet final. Penndorf's calculation is heavily critized by Moses and Wu. But these authors consider only the contribution of oxygen to the energy balance which determines the temperature. It seems doubtful whether this is justified on account of the radiation absorbed by nitrogen.

The theoretical approaches made hitherto are based only on static considerations, neglecting completely dynamical influences. It is now well known that turbulent movements and ionospheric winds are rather important in the *E* layer. In that case the thickness of the transition layer, obtained by static calculations only, cannot be identified with the observed parameters of a real layer.[11] We can suppose that even the sharp theoretical transition layers calculated by Penndorf (Fig. 1) are extended by turbulence. Thus, without judging the different hypotheses, we think that the numerical results of Moses and Wu are approximately acceptable. Recent rocket observations at 1475A show that absorption begins at 140 km and reaches a maximum at 105 km.[12] As this wavelength is absorbed by O_2, this result must be interpreted in favor of a broad transition layer situated between 90 and 150 km.

[1] D. R. Bates and F. Hoyle, Terr. Mag. **53**, 51 (1948).
[2] E. Argence and K. Rawer, Compt. rend. **233**, 1208 (1951).
[3] G. Elwert, Z. Naturforsch. **7a**, 202 (1952).
[4] M. Nicolet, Mem. Roy. Met. Inst. Belgium **19**, 124 (1945).
[5] D. C. Choudhury, Phys. Rev. **88**, 405 (1952).
[6] E. Argence and K. Rawer, Ann. géophys. **9**, 1 (1953).
[7] R. Penndorf, J. Geophys. Research **54**, 7 (1949).
[8] R. C. Majumdar, Indian J. Phys. **12**, 75 (1938).
[9] H. Rakshit, Indian J. Phys. **21**, 57 (1947).
[10] H. E. Moses and T. Y. Wu, Phys. Rev. **83**, 109 (1951).
[11] M. Nicolet (private communication).
[12] Friedman, Lichtman, and Bryan, Phys. Rev. **83**, 1025 (1951).

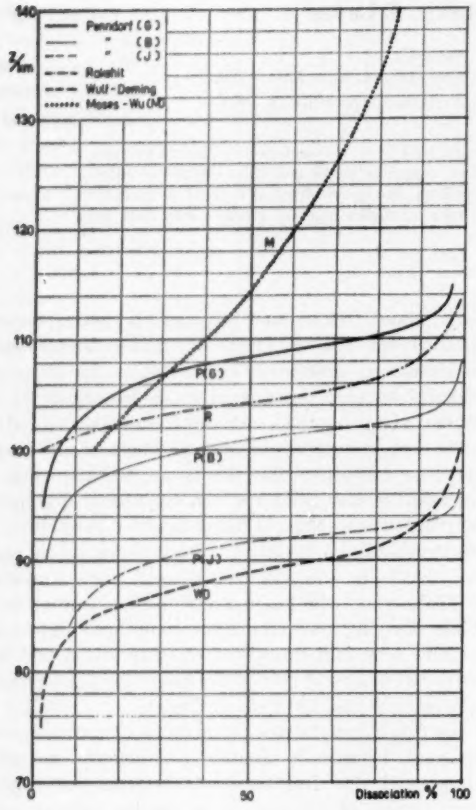

FIG. 1. Rate of dissociation of oxygen molecules in the high atmosphere.

Distribution of Ionization

We have calculated the ionization obtained in the case of a monochromatic radiation ionizing only O_2 molecules for some models of transition layers.[6] The distribution of the number density N' of O_2 is supposed to be

$$N' = N_0'(1-b_z')^{1/b'h_0}$$

above the lower transition limit, and a constant fraction of the total number density below it. The data of our four models *A*, *B*, *C*, *D* have been indicated in Fig. 2 together with Penndorf's theoretical density values and that of Moses and Wu. The corresponding ionization was calculated with different assumed absorption coefficients (*viz.*, cross sections) such that in each case the maximum of ionization was found at 125 km (this height is well known from ionospheric observations). Thus we obtain the ionization curves of Fig. 3. Curve *C* and *D* must be rejected because the calculated thickness is too small; curve *B* is most similar to the result of ionospheric observations on the *E* layer. The cross section effective for ionization should be 1.33×10^{-17} cm^2 in this case. Choudhury used a simple exponential variation of N' as an approximation to the values of Moses and Wu. He found a cross section of 0.66–0.88 $\times10^{-17}$ cm^2; the maximum of ionization was assumed to be at 120 km. Taking account of the difference in altitude, the two results agree very well.

The corresponding model *B* for the transition layer (Fig. 2) is rather broad and high. In any case the transition should be found at high altitudes, if the usual theory of specific ionization could be acceptable at all.

Solar Radiation Causing Ionization

The intensity of solar radiation producing the *E* layer can be obtained from the observed electron density and the cross section calculated above; it should be 9.3×10^7 quanta cm^{-2} sec^{-1}. In the range of preionization, 900 to 1020A, the intensity in the solar spectrum at the top of the atmosphere is estimated to be several times this number.[13] Unfortunately, rocket measurements have not yet been made in this special wavelength range. Now Elwert[14] doubts whether this radiation could arrive at the altitude of 130 km without heavy losses by N_2 absorption. Recently published results on N_2 absorption[15] indicate a high absorption

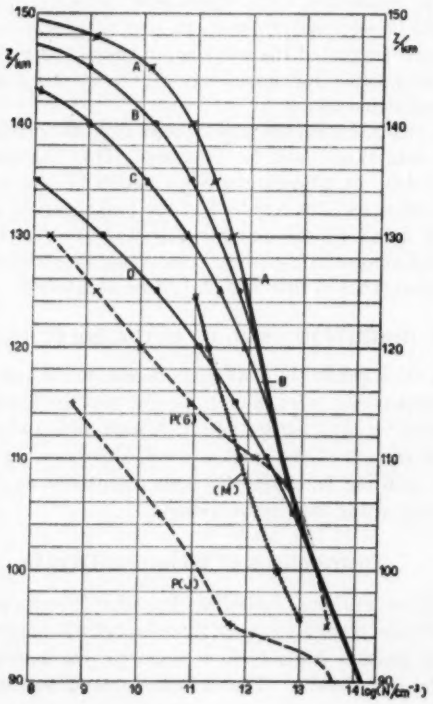

FIG. 2. Particle density of molecular oxygen (broken lines: results of Penndorf and Moses and Wu; full lines: our models *A*, *B*, *C*, *D*).

[13] M. Nicolet, Ann. géophys. 8, 141 (1952).
[14] G. Elwert (private communication).
[15] Weissler, Lee, and Mohr, J. Opt. Soc. Am. 42, 84 (1952).

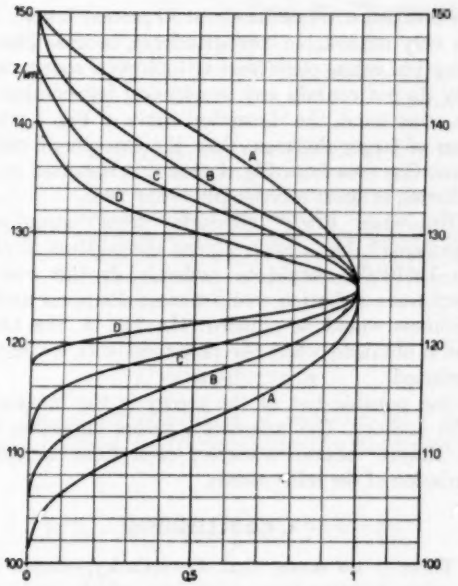

FIG. 3. Calculated electron density for our four models (the recombination coefficient has been assumed to be constant; the maximum electron density is the unit on the abscissa).

below 980A; the absorption is $\exp(-500z/\text{cm})$ for standard conditions. This radiation cannot reach the height of the *E* region. Above 980A there seems to exist a "window"; the absorption coefficient has a minimum at 1026A with 60 instead of 500 cm^{-1}. With this value we find an attenuation of $1/e$ at a pressure of 1.6×10^{-5}mm Hg for the case of absorption in the atmosphere.[16] This pressure exists at an altitude of about 130 km.[17,18] Only radiation from the "window" could thus reach the *E* region without a too heavy loss. It is questionable whether the amount of incoming radiation in the window is sufficient. As L_γ(972A) is outside the window, and as the quantum energy of L_β(1025A) is probably insufficient to produce preionization, Elwert's objection could be justified. Only if there were another important radiation between 980 and 1020A could the theory of preionization be maintained.

3. IONIZATION BY X-RAYS

In this case the ionization holds for all atmospheric molecules; an estimate of its effective section can be obtained with a simple model calculation. Instead of the exponential density variation used by Choudhury, we prefer a model with a linear variation of temperature.[19] The height of the maximum electron density, depends on the recombination law. We think there are arguments that it is nearer to a variation of α (the recombination coefficient) with the square root of number density[20] than to a constant value α. The condition for the maximum is[6]

$$\sigma \sec\chi N_m H_m/(1+bH_m)=\tfrac{1}{2} \quad \text{or} \quad 1,$$

for these two assumptions, where σ is the cross section, χ the zenith angle, N_m, H_m the number density and scale height at the height of the maximum, b the coefficient of the linear temperature gradient; $b\approx0.015\ \text{km}^{-1}$. It follows for a maximum at 125 km that

$$\sigma \sec\chi=0.68 \quad \text{or} \quad 1.35\times10^{-18}\ \text{cm}^2.$$

This is lower than Choudhury's estimate.[5] Consequently we think that a longer wavelength of radiation must be considered, namely 100···200A (60···120 ev, instead of 181).[21]

Elwert's calculation of solar x-ray emission[3] is based exclusively on astrophysical data. He only introduced

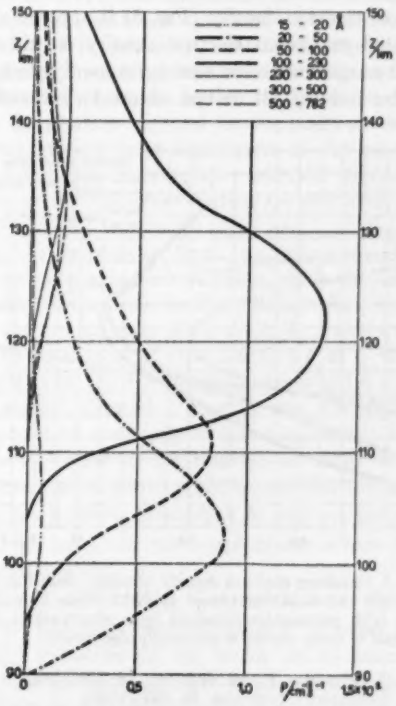

FIG. 4. Electron production for different wavelength ranges in the case of Elwert's corona model with 10^6 °K.

[16] This N_2 absorption will not produce ionization because the quantum energy of 12···13 ev is lower than the ionization potential of N_2.

[17] Havens, Koll, and La Gow, J. Geophys. Research **57**, 59 (1952).

[18] An eventual dissociation of N_2 does not seriously destroy this argument because it is certainly small at altitudes inferior to 150 km.

[19] J. A. Gledhill and M. E. Szendrei, Proc. Phys. Soc. (London) **B63**, 427 (1950); M. Nicolet and L. Bossy, Ann. géophys. **5**, 275 (1949).

[20] Argence, Mayot, and Rawer, Ann. géophys. **6**, 242 (1950).

[21] The rocket measurements of Friedman[12] concern only wavelengths below 10A; the maximum of absorption is obtained for 87 km. This radiation could thus be perhaps one cause of the ionization of the *D* layer, but not of the *E* layer.

the electron density of the solar corona, its temperature, and the cosmical relative abundance of elements. Taking account of ionization, recombination, and excitation he finds radiation by free-free, free-bound, and bound-bound transitions. The most important contribution comes from the latter; the most probable excitation process is by electron impact.

With Elwert's radiation values[3] obtained for a temperature of 10^6 °K and with the values of the absorption cross section given by Schneider[22] and Siedentopf,[23] we have calculated the absorption for different ranges of wavelength as a function of the height (Fig. 4.) The abcissas of the curves are proportional to the production rate of electrons. Summing up the different contributions, one obtains the total amount of production. From this, with assuming the coefficient of effective recombination to vary with the altitude,[24] we obtained the distribution of electron density (Fig. 5). We have assumed a high efficiency of ionization, only 15.8 ev for each ionization; perhaps this value should be as high as 30 ev,[25] but there is little information about very soft x-rays. From comparison with experimental curves of electron density (Fig. 5) it appears that the calculated maximum electron density is midway beween the experimental values corresponding to high and low solar activity. (If we had assumed a lower efficiency of ionization it would be about 30 percent lower). This is a very remarkable correspondence, because Elwert's numerical values come from astrophysical sources only; they do not contain any geophysical information. On the other hand, the theoretical curve of Fig. 5 gives a layer of larger thickness than the ionospheric records show. One sees from Fig. 4 that this is caused by the radiation of short wavelength, below 50A.

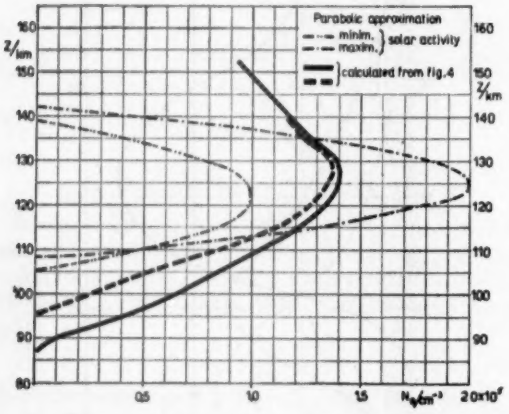

FIG. 5. Resulting electron density obtained from Fig. 4 (broken line, when the radiation ceases at 50A). Thin lines: parabolic models with parameters obtained from observations (only the lower half of these curves is physically significant).

Meanwhile Elwert concluded from astrophysical arguments[26] that a lower corona temperature of about $6\cdots7\times10^5$ °K should be probable. In this case the short wave radiation would almost disappear and the thickness would be reduced. (In Fig. 5, the broken line is obtained when the radiations with $\lambda<50$A are neglected).

One possible test of the theory is the influence of solar activity. This influence is rather important (Fig. 5). Elwert[26] obtains a correct value from the density variation of the solar corona.

4. CONCLUSIONS

There is no doubt that Elwert's hypothesis is attractive. Starting from solar data, which seems to be rather well established, he calculates almost exactly the shape and intensity of *E* layer ionization. It seems highly probable that the most important part of this ionization comes from x-rays.[27]

On the other hand, dissociation of oxygen certainly exists, and absorption data do not completely exclude ultraviolet radiation of 1000A from reaching the altitude of 120 km. Perhaps an extra radiation of this wavelength could be responsible for the stratifications of the *E* layer, like *E*2. But there is yet another unsolved question for this hypothesis of preionization, i.e., whether its effective cross section is sufficiently high.

We agree with Choudhury, that perhaps two different ionization processes could exist for the altitude of the *E* layer. Some sunrise effects could be better understood with this assumption.[28] However, contrary to this author's opinion, we would consider ionization by x-rays as the more important process, controlling the normal features of the *E* layer.

It seems that the rather complicated structure of ionization in the *D* and *E* regions must be attributed to several different absorption processes. As it is now known that turbulence and winds are normal features at the height of the *E* region, future theories of layer formation must take account of this influence.

[22] E. G. Schneider, J. Opt. Soc. Am. **30**, 128 (1940).
[23] H. Siedentopf, Naturwiss. **35**, 289 (1948).
[24] The variation is obtained from experimental values of different authors; in the range 80···310 km the interpolation formula is $\log\alpha/\text{cm}^3\ \text{sec}^{-1}=-8.4+0.37\zeta+0.68\zeta^3$, $\zeta=(190\text{ km}-Z)/100$ km. Details will be published later.
[25] H. Kulenkampff, Ann. Physik **79**, 97 (1926).
[26] G. Elwert, Z. Naturforsch. (to be published).
[27] In a recently published spectroscopic work on nitrogen ions, M. Dufay comes to similar conclusions [Ann. phys. **8**, 813 (1953)].
[28] E. Theissen, Naturwiss. **34**, 371 (1947); R. Lindquist, J. Geophys. Research **57**, 439 (1952).

Density Fluctuations at Low Temperatures

P. J. PRICE*
Institute for Advanced Study, Princeton, New Jersey
(Received August 20, 1953)

The applicability to a quantum liquid of the standard classical formula connecting the compressibility with the coherent scattering cross section for large wavelengths, questioned by the author in a previous paper, is examined. The correctness of the standard formula is proved (*a*) at absolute zero (the density fluctuations being infranormal); (*b*) under quantum conditions for all temperatures at which the Wigner expansion converges (it is conjectured that for liquid helium the expansion may diverge below the lambda-point); and (*c*) for a one-dimensional crystal for all temperatures. These results, while they stop short of a complete proof of the standard classical formula for all conditions, do extend considerably our knowledge of its range of validity.

1. INTRODUCTION

IN the coherent scattering of radiation by a monatomic liquid, the scattered intensity is proportional to[1]

$$R(|\mathbf{k}_1-\mathbf{k}_2|)=1+\rho\int h(r)e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{r}}d^3\mathbf{r}, \tag{1}$$

where ρ is the number density of atoms, supposed uniform over the scattering volume, $1+h(r)=g(r)$ is the normalized radial distribution function, and $\mathbf{k}_1$ and $\mathbf{k}_2$ are the wave vectors in the incident and scattered directions. Consequently, for large wavelengths the atomic scattering cross section is multiplied by an interference factor

$$R(0)=1+\rho\int_0^\infty h(r)4\pi r^2dr. \tag{2}$$

This factor is proportional to the mean square fluctuation of the number of atoms in a fixed region, or "cell," in the liquid, which is given by

$$\langle M^2\rangle-\langle M\rangle^2=\langle M\rangle\left[1+\rho\int_0^\infty h\,4\pi r^2dr\right], \tag{3}$$

where M is the number of atoms in the cell in a single observation, the bracket $\langle\ \rangle$ denotes an average over a large number of observations, $\langle M\rangle$ is small compared with the total number N of atoms in the assembly, and the linear dimensions of the cell are large compared with the values of a for which $\int_a^\infty r^2hdr$ is appreciable compared with $\int_0^\infty r^2hdr$. By combining (2) and (3) with the Gibbs fluctuation formula

$$\langle M^2\rangle-\langle M\rangle^2=\langle M\rangle\rho\kappa T\chi_T, \tag{4}$$

where χ_T is the isothermal compressibility and κ is Boltzmann's constant, one obtains the standard relation connecting the scattering factor R with thermodynamic magnitudes:

$$R(0)=\rho\kappa T\chi_T. \tag{5}$$

In a previous paper[2] (to be referred to here as I) the formula (5) was called in question for the conditions at low temperatures where quantum physics determines the state of the assembly, such as in liquid helium. Formulas (1) and (2), which are derived from a purely geometrical analysis in terms of the reduced probability distribution $n_2(\mathbf{r}_1,\mathbf{r}_2)$ for the positions of a pair of atoms, remain valid irrespective of the mechanism determining that distribution—whether thermal fluctuations or the zero-point motion are dominant. It was argued in I that (4), on the other hand, cannot remain valid as the temperature tends to absolute zero, because the right-hand side then vanishes, while the left-hand side, which then measures the quantum fluctuations of M for the ground state of the assembly, cannot be zero since the observable M (the function F of I) does not commute with the Hamiltonian and is not a constant of the motion. (Stated in another way, the fluctuation of M could be zero only if the distribution function $n_N(\mathbf{r}_1\cdots\mathbf{r}_N)$ were a combination of delta functions giving every atom a fixed position; and this is impossible by the uncertainty principle.) It was suggested that (4) and (5) hold only for temperatures where classical mechanics applies.

Further consideration has shown that the above argument about the limit of (4) as $T\to0$ is not a valid one. The vanishing of the factor $\rho\kappa T\chi_T$ may imply merely that the fluctuations in M are *infranormal*: that $\langle M^2\rangle-\langle M\rangle^2$ does not increase "as fast as $\langle M\rangle$" when the volume of the cell, $v=\langle M\rangle/\rho$, increases (together with N) indefinitely.[3] For example, the finite size of a (still large) cell may be taken into account by an extension

* Now at Watson Scientific Computing Laboratory, Columbia University, New York, New York.

[1] For derivations of formulas (1)–(5), see J. de Boer, Repts. Progr. in Phys. **12**, 305 (1949), Sec. 10; H. S. Green, *The Molecular Theory of Fluids* (Interscience Publishers, New York, 1952), Secs. 3.1, 3.2; J. Yvon, *Fluctuations en Densité*, Actualités Scientifiques et Industrielles, No. 542 (Hermann, Paris, 1937), Chap. 1.

[2] P. J. Price, Phys. Rev. **86**, 495 (1952).

[3] I am indebted to G. Placzek and L. van Hove for drawing my attention to this possibility and its relevance, and for championing the universal validity of (4).

of (3) to

$$\langle M^2\rangle-\langle M\rangle^2=v\cdot\rho\left[1+\rho\int_0^\infty 4\pi r^2 h\,dr\right]$$
$$+A\cdot\rho^2\int_0^\infty ds\int_s^\infty 2\pi r(s-r)h\,dr, \quad (3')$$

where A is the surface area of the cell, provided the integrals converge. The coefficient of v may then vanish while the fluctuation of M remains finite, being in the limit of large v proportional to A. However, it remained true that, more generally, (4) *had not been validly proved* except "for classical conditions." If it should nevertheless be correct in general (giving the coefficient of the asymptotic normal part of the fluctuation of M), such a simple connection between the temperature and thermodynamic compressibility, on the one hand, and the quantum fluctuations of M for pure states of the assembly, seemed physically mysterious. It therefore seemed worthwhile to investigate whether (4), and hence (5), could be satisfactorily proved (or disproved) for quantum conditions.

In this paper it is shown that (4) holds certainly for a part, and possibly for the whole, of the range of temperature in which quantum mechanics applies. In Sec. 2 it is proved that the expression (2), $R(0)$, is in fact zero for the ground state, so that (4) and (5) hold at absolute zero (the fluctuation of M being infranormal). In Sec. 3 it is shown that (4) must hold in quantum conditions at all temperatures for which the Wigner expansion, as defined there, converges. The interpretation of these results for liquid helium is discussed. In Sec. 4 we consider a condensed system for which a complete exact solution is possible, namely a one-dimensional *crystal*, and prove that the equivalent of (4) is correct for all temperatures. For absolute zero, the explicit infranormal dependence on $\langle M\rangle$ of the fluctuation of M is obtained. These results stop short of a proof of (4) for all conditions, though they extend considerably our knowledge of its range of validity.

2. ABSOLUTE ZERO

We shall prove that for the ground state of a liquid $R(k)\leqslant$ (const.)k,[4] and hence that $R(0)$ is zero. Let the normalized wave functions of the stationary states of the system be $\psi_0, \psi_1, \cdots, \psi_\nu, \cdots$, with energies $E_0, E_1, \cdots, E_\nu, \cdots$, where ψ_0 and E_0 refer to the ground state. Let

$$\phi=\sum_{s=1}^{N}\cos(\mathbf{k}\cdot\mathbf{r}_s), \quad (6)$$

and

$$\phi\psi_0=\psi=\sum_{\nu=0}^{\infty}a_\nu\psi_\nu. \quad (7)$$

[4] This inequality, and the outline of the derivation presented here, are due to L. Onsager (private communication, 1953).

Then it follows at once that[5]

$$(\psi|\psi)=(\psi_0|\phi^2|\psi_0)=\tfrac{1}{2}S^2R(k),$$

or

$$\textstyle\sum_\nu|a_\nu|^2=\tfrac{1}{2}S^2R(k). \quad (8)$$

Secondly we note that, with the Hamiltonian of the system of form

$$H=-(\hbar^2/2m)\textstyle\sum_s\nabla_s^2+\Phi(\mathbf{r}_1,\mathbf{r}_2\cdots\mathbf{r}_N) \quad (9)$$

(where Φ may depend also on spin coordinates), the expectation $(\psi|H-E_0|\psi)$ may be transformed, by subtracting a vanishing surface integral, to

$$(\hbar^2/2m)(\psi_0|[\textstyle\sum_s|\nabla_s\phi|^2]|\psi_0),$$

and hence to $(\hbar^2/2m)\frac{1}{2}Nk^2$. Thus,

$$\textstyle\sum_\nu|a_\nu|^2(E_\nu-E_0)=(\hbar^2/2m)\frac{1}{2}Nk^2. \quad (10)$$

To obtain the third relation which is required together with (8) and (10), we imagine the system perturbed by the addition of a *small* potential $\Omega\phi$, where ϕ is given by (6) and Ω is a constant. The Hamiltonian of the perturbed system is then $H'=H+\Omega\phi$. An "ordinary" liquid behaves like an ideal elastic fluid in that the change in the internal energy $E=\langle H\rangle$, on perturbation by a potential field, may be expressed in terms of the redistribution of the number density ρ:

$$E-E_0=\int[\epsilon(\rho')-\epsilon(\rho)]d^3\mathbf{r}$$
$$=\tfrac{1}{2}\epsilon''(\rho)\int(\rho'-\rho)^2d^3\mathbf{r}+\cdots, \quad (11)$$

and hence the *total* perturbation energy in this case—the minimum of $E-E_0+\Omega\int\rho\cos(\mathbf{k}\cdot\mathbf{r})d^3\mathbf{r}$—would be:

$$\text{perturbation energy}=[V/4\epsilon''(\rho)]\Omega^2+O(\Omega^3), \quad (12)$$

where $V=N/\rho$ is the volume occupied by the assembly. A justification of (12) for the ground state of a quantum liquid is given in the appendix. Then, comparing (12) with the standard formula of second-order perturbation theory, we have

$$\textstyle\sum_\nu{}^{(0)}|a_\nu|^2/(E_\nu-E_0)=\frac{1}{4}N/\rho\epsilon''(\rho), \quad (13)$$

where the sum $\sum_\nu{}^{(0)}$ excludes $\nu=0$.

Now,

$$\textstyle\sum_\mu{}^{(0)}|a_\mu|^2(E_\mu-E_0)\sum_\nu{}^{(0)}|a_\nu|^2/(E_\nu-E_0)$$
$$=\textstyle\sum_\mu{}^{(0)}\sum_\nu{}^{(0)}|a_\mu|^2|a_\nu|^2\left[1+\dfrac{(E_\mu-E_\nu)^2}{2(E_\mu-E_0)(E_\nu-E_0)}\right]$$
$$\geqslant(\textstyle\sum_\nu{}^{(0)}|a_\nu|^2)^2.$$

[5] The relation (8), and hence the inequality (15), is actually true only when $\mathbf{k}$ takes one of the "spectrum values," which are those for which the integral of ϕ over the volume of the ensemble vanishes. However, the spectrum values may be made as close as we please by choosing N large enough (at constant ρ); and $R(0)$, given by (2), is to be understood as the limit of $R(k)$ as $\mathbf{k}$ tends to zero *through spectrum values*.

With $\mathbf{k}$ one of the spectrum values,[5] $a_0=0$ and hence we have

$$\sum_\mu |a_\mu|^2(E_\mu-E_0)\sum_\nu{}^{(0)}|a_\nu|^2/(E_\nu-E_0) \geqslant [\sum_\mu |a_\mu|^2]^2. \quad (14)$$

Substituting (8), (10) and (13) into (14), we obtain Onsager's inequality:

$$R(k) \leqslant \hbar(4m\rho\epsilon'')^{-\frac{1}{2}}k, \quad (15)$$

for the ground state. In terms of the calculated velocity of sound c, given by

$$mc^2=(\chi\rho)^{-1}=\rho\epsilon'',$$

(15) may be written

$$R(k) \leqslant (\hbar/2mc)k. \quad (15')$$

It follows from (15) that $R(0)$—the coefficient of the normal density fluctuations, and the scattering factor for long wavelengths—vanishes at absolute zero. The above proof is independent of the symmetry—Bose, Fermi or classical Boltzmann—of the wave functions (which simply determines the definition of the $\{\psi_\nu\}$), and of the spin of the atoms. It clearly applies equally to one (monatomic) component of a mixture.

3. FINITE TEMPERATURES: THE WIGNER SERIES

In this section we set out the proof of (4) for a *classical* liquid, and then examine how far this proof may be extended to temperatures at which quantum mechanics must apply. For a classical liquid, the probability distribution of atomic configurations $(\mathbf{r}_1,\cdots\mathbf{r}_N)$ is

$$n_N(\mathbf{r}_1,\cdots\mathbf{r}_N)=(\lambda^N/N!)\exp[(F-\Phi)/\kappa T], \quad (16)$$

where Φ is the interaction potential and F the free energy, and λ is a function of the temperature T only. The free energy is accordingly given by

$$N!\exp(-F/\kappa T) =\lambda^N\int d^3\mathbf{r}_1\cdots\int d^3\mathbf{r}_N\exp(-\Phi/\kappa T), \quad (17)$$

where the N integrations are each taken over the volume of the assembly. For N large, F is given by

$$F(N,V,T)=Nf(N/V,T). \quad (18)$$

The probability of finding M atoms in a cell A of volume v, and $N-M$ atoms in the remainder B of the space, is

$$w(M)=[N!/M!(N-M)!] \times\int^A d^3\mathbf{r}_1\cdots\int^A d^3\mathbf{r}_M\int^B d^3\mathbf{r}_{M+1}\cdots\int^B d^3\mathbf{r}_N\, n_N,$$

or, by (16),

$$w=[\lambda^M/M!][\lambda^{N-M}/(N-M)!] \times\int^A d^3\mathbf{r}_1\cdots\int^B d^3\mathbf{r}_N\exp[(F-\Phi)/\kappa T]. \quad (19)$$

Now, if we have

$$\Phi=\sum_{s<t}\sum\phi(r_{st}),\quad \mathbf{r}_{st}=\mathbf{r}_s-\mathbf{r}_t, \quad (20)$$

and if $\phi(r)$ is effectively zero outside a definite sphere of influence, then it is obvious that the terms

$$\Phi_{AB}=\sum_{s=1}^{M}\sum_{t=M+1}^{N}\phi(r_{st})$$

may be omitted from $\Phi=\Phi_A+\Phi_B+\Phi_{AB}$ in (19) without appreciably affecting the value of the integral, provided M and $N-M$ are large numbers. Then, on dropping Φ_{AB} from Φ in (19), the integrals over A and over B are two independent factors and by (17) and (18)

$$w\simeq\exp\left\{\left[Nf\left(\frac{N}{V},T\right)-Mf\left(\frac{M}{v},T\right) -(N-M)f\left(\frac{N-M}{V-v},T\right)\right]\Big/\kappa T\right\}. \quad (21)$$

This expression (21) has a maximum for $M=vN/V=M_0$. Then for $M-M_0\ll M_0$ we may expand the exponent in powers of $M-M_0$ as far as the square only, and find

$$w\simeq\exp[-\alpha(M-M_0)^2],\quad \alpha=V/2\kappa T\chi_T M_0(N-M_0). \quad (21')$$

Hence, with $M_0\ll N$, we obtain (4). The addition of three-atom, four-atom, etc., terms to the two-atom potential (20) does not affect the above argument, provided the terms Φ_{AB}, consisting of the contribution "acting across the cell boundary" of interactions linking atoms $1\cdots M$ in A to $M+1\cdots N$ in B, may still be neglected in evaluating (19). Obviously, if Φ is composed of various n-atom interactions of type $\sum\phi(\mathbf{r}_1,\cdots,\mathbf{r}_n)$, only up to some definite degree $q\geqslant n$, then this condition may be fulfilled merely by choosing v large enough (with v/V kept small) for the mean number of atoms in the cell, $M_0=\rho v$, to be large compared with q^3.

The classical distribution (16) is of course correct only at high enough temperatures, but the validity of the above argument may be extended to lower temperatures by expressing the distribution given by quantum mechanics as a series in ascending powers of Planck's constant $2\pi\hbar$.[6] Let us do this by replacing Φ

[6] E. P. Wigner, Phys. Rev. **40**, 749 (1932); M. L. Goldberger and E. N. Adams II, J. Chem. Phys. **20**, 240 (1952).

in (16) by

$$\Phi^* = \Phi + \sum_{n=1}^{\infty} (h^2/2m\kappa T)^n \theta_n(\mathbf{r}_1, \cdots \mathbf{r}_N; T), \quad (22)$$

where θ_n is given by a sum over multiple interactions up to, say, q_n atomic. The first of the θ_n is given[7] by

$$6\theta_1 = \sum_s [\nabla_s^2 \Phi - (1/2\kappa T) |\nabla_s \Phi|^2]. \quad (23)$$

If Φ is given by (20) (pair interactions), then (23) becomes

$$6\theta_1 = \sum_{s<t} \{\phi''(r_{st}) + (2/r_{st})\phi'(r_{rs}) - (1/\kappa T)[\phi'(r_{st})]^2\}$$

$$- (1/2\kappa T) \sum_{s<t\neq u} \sum (\mathbf{r}_{ts} \cdot \mathbf{r}_{tu}/r_{ts} r_{tu}) \phi'(r_{ts}) \phi'(r_{tu}), \quad (23')$$

where $\mathbf{r}_{ts} \cdot \mathbf{r}_{tu}$ means $x_{ts}x_{tu} + y_{ts}y_{tu} + z_{ts}z_{tu}$. Thus, with pair interactions ($q=2$) for the actual potential Φ, we have $q_1 = 3$ (and, in fact, $q_2 = 6$). More generally, q_n increases monotonically with n. If, therefore, the series (22) may in practice be terminated at some point —say at θ_r, so that Φ^* contains "cluster" functions up to q_r atomic—then the previous argument leading to (21′) still applies and (4) is still verified. We may expect, however, that as quantum effects become more important (at lower temperatures) the series (22) has to be taken to an increasing number of terms in evaluating (19) with sufficient accuracy, and hence progressively higher cluster-numbers are involved. (Looked at a little differently: if the terms of Φ^* were regrouped in inverse powers of κT, we would expect the coefficient of $(\kappa T)^{-n}$ to contain cluster functions up to $q = n+2$; and thus with decreasing temperature higher clusters should have to be taken into account.) Thus, if at some temperature the series (22) ceases to converge as a functional of Φ [which could happen without the onset of a discontinuity in F, as calculated by (17) in terms of Φ^*] then the above argument for (4), based on restricting Φ^* to clusters of not more than some definite degree q, ceases to be valid at that point. The argument of this section furnishes a proof of (4) under those conditions where the Wigner series, in the form (22), converges.

It might be conjectured that in the case of liquid helium (He^4) the lambda transition, where there are singularities in the second derivatives of F, marks the position of a divergence of the Wigner expansion. (There is some recent experimental evidence,[8] which is of interest in this connection, that the effective range of interaction between atoms through Φ^* increases very considerably from 2.5° to the lambda point.) In this case, the above argument does not exclude the possibility that there is a departure from (4) and (5) between the lambda point and (but not including) absolute zero. On the other hand, if the Wigner series (22) does converge for all temperatures then the above argument shows that (4) and (5) hold for all temperatures.

[7] Equation (23) refers to the series for classical (Boltzmann) statistics; but similar series may be established for the Bose and Fermi cases [see H. S. Green, J. Chem. Phys. 19, 955 (1951)], and the argument which follows covers all three.

[8] D. F. Brewer and K. Mendelssohn, Phil. Mag. 44, 559 (1953).

4. ONE-DIMENSIONAL CRYSTAL

In this section we take the case of a one-dimensional "crystal," a linear chain of atoms with nearest-neighbor harmonic interactions, and show that the analog, for this case, of (4) holds *for all temperatures*. Let the number of atoms in the chain, each of mass m, be N, and let their static equilibrium spacing be a. Let the coordinate of the rth atom in the chain be

$$z_r = ra + x_r, \quad (24)$$

so that x_r is the displacement from equilibrium. We shall determine, instead of the fluctuation in the linear density of atoms (which does not have a convenient definition for a lattice), the equivalent mean square deviation of the distance between two atoms,

$$\Delta_n = \langle [(z_{n+r} - z_r) - \langle z_{n+r} - z_r \rangle]^2 \rangle$$
$$= \langle (x_{n+r} - x_r)^2 \rangle, \quad (25)$$

as a function of $\langle (z_{n+r} - z_r) \rangle = na$. It is convenient to adopt cyclic boundary conditions, treating the first atom as the $(N+1)$th. Then the Hamiltonian of the assembly is

$$H = \sum_{r=1}^{N} [p_r^2/2m + \tfrac{1}{2} m\omega_0^2 (x_{r+1} - x_r)^2], \quad (26)$$

where the p_r are the momenta conjugate to the x_r. Taking N as even for convenience, the Hamiltonian is made separable by the transformation to normal coordinates:[9]

$$x_r = (2/N)^{\frac{1}{2}} \sum_{s=1}^{N/2} [u_s \sin(rs\theta) + v_s \cos(rs\theta)],$$

$$p_r = (2/N)^{\frac{1}{2}} \sum_{s=1}^{N/2} [\sigma_s \sin(rs\theta) + \tau_s \cos(rs\theta)], \quad (27)$$

where $\theta = 2\pi/N$. σ_s and τ_s are the momenta conjugate to the normal coordinates u_s, v_s, and fulfill with them the standard commutation relations. Then H transforms to

$$H = \sum_s [(\sigma_s^2/2m + \tfrac{1}{2} m\omega_s^2 u_s^2) + (\tau_s^2/2m + \tfrac{1}{2} m\omega_s^2 v_s^2)]$$
$$= \sum_s [H_s^{(1)} + H_s^{(2)}], \quad (28)$$

where

$$\omega_s = 2\omega_0 \sin(\tfrac{1}{2} s\theta). \quad (29)$$

From (29) the velocity of long waves, under the conditions of classical mechanics, is $c_0 = a\omega_0$. Each of the normal coordinates thus represents an harmonic oscillator; and hence, for a stationary state of the ensemble,

[9] The center-of-gravity coordinate corresponding to $s=0$ is of no interest in this analysis and may be ignored.

or for any equilibrium state,

$$\langle u_s v_t\rangle = 0,$$
$$\langle u_s u_t\rangle = \langle H_s^{(1)}\rangle \delta_{st}/m\omega_s^2, \tag{30}$$
$$\langle v_s v_t\rangle = \langle H_s^{(2)}\rangle \delta_{st}/m\omega_s^2.$$

From (27) and (30),

$$\langle x_r x_{n+r}\rangle = (2/N)\textstyle\sum_s \frac{1}{2}(\langle u_s^2\rangle + \langle v_s^2\rangle)\cos(ns\theta),$$
$$\langle (x_{n+r}-x_r)^2\rangle = (2/N)\textstyle\sum_s 4\sin^2(\frac{1}{2}ns\theta)\frac{1}{2}(\langle u_s^2\rangle + \langle v_s^2\rangle),$$
$$\Delta_n = (2/N)\textstyle\sum_s [\sin^2(\frac{1}{2}ns\theta)/m^2\omega_0^2\sin^2(\frac{1}{2}s\theta)] \times \frac{1}{2}(\langle H_s^{(1)}\rangle + \langle H_s^{(2)}\rangle). \tag{31}$$

For the ground state of the assembly,

$$\langle H_s^{(1)}\rangle = \langle H_s^{(2)}\rangle = \tfrac{1}{2}\hbar\omega_s = \hbar\omega_0 \sin(\tfrac{1}{2}s\theta),$$

and hence

$$\Delta_n = (\hbar/m\omega_0)(2/N)\sum_{s=1}^{N/2}[\sin^2(\tfrac{1}{2}ns\theta)/\sin(\tfrac{1}{2}s\theta)]. \tag{32}$$

If N is large, and $n \ll N$, we may replace the sum (32) by an integral over $\phi = \frac{1}{2}s\theta$:

$$\Delta_n = \left(\frac{\hbar}{m\omega_0}\right)\frac{2}{\pi}\int_0^{\pi/2}\frac{\sin^2(n\phi)d\phi}{\sin\phi}. \tag{32'}$$

The integral in (32′) is

$$\sum_{p=1}^{n}(2p-1)^{-1} \simeq \text{const.} + \tfrac{1}{2}\log n + O(1/n).$$

Hence $\Delta_n/an \to 0$ as $n \to \infty$, and *the fluctuation is infranormal*. Although Δ_n tends to infinity when an does, it does not increase "as fast as n" but only "as fast as $\log n$."

For thermal equilibrium at a temperature T, we have for the time average of the expectation energy of an oscillator:

$$\langle H_s^{(1)}\rangle = \langle H_s^{(2)}\rangle = \kappa T\epsilon\coth\epsilon \tag{33}$$

where $\epsilon = \hbar\omega_s/2\kappa T = (\hbar\omega_0/\kappa T)\sin(\frac{1}{2}s\theta)$. Making the same transformation as from (32) to (32′),

$$\Delta_n = \left(\frac{\kappa T}{m\omega_0^2}\right)\frac{2}{\pi}\int_0^{\pi/2}G(\phi)\left[\frac{\sin(n\phi)}{\sin\phi}\right]^2 d\phi, \tag{34}$$

where

$$G(\phi) = [\hbar\omega_0\sin\phi/\kappa T]\coth[\hbar\omega_0\sin\phi/\kappa T].$$

The form of the integral in (34),

$$I_n(F) = \int_0^{\pi/2}F(x)[\sin(nx)/\sin x]^2 dx,$$

is essentially Fejér's integral,[10] and has the property

$$\operatorname*{Lim}_{n\to\infty}(I/n) = (\pi/2)\operatorname*{Lim}_{x\to+0}F(x),$$

provided $F(x)$ is integrable in $(0,\pi/2)$. In the present case $F(+0) = G(0) = 1$, and hence

$$\operatorname*{Lim}_{n\to\infty}(\Delta_n/an) = \kappa T/am\omega_0^2 \tag{35}$$

for all T. To relate (35) to the one-dimensional compressibility, we have to regard a as a variable. The expectation of H (26) is a function of ω_0, T and N, and does not depend on a; but we should add to the potential energy in (26) a term

$$V_0 = \tfrac{1}{2}m\omega_0^2 a^2 N \tag{36}$$

to take account of the change in the total energy, E, on compression of the lattice. Then $(\partial^2 E/\partial a^2)_{N,T} = m\omega_0^2 N$. The isothermal compressibility is thus

$$\chi_T = aN/[a^2(\partial^2 E/\partial a^2)_{N,T}] = 1/am\omega_0^2; \tag{37}$$

and (35) may be written

$$\Delta_n \simeq (a^2 n)\rho\kappa T\chi_T, \quad 1 \ll n \ll N, \tag{38}$$

where $\rho = 1/a$ is the linear density of atoms. Formula (38) is the precise analog of (4), in terms of the *inverse* of a mean linear density $n/(z_{n+r}-z_r)$, and has been found here to be correct for all temperatures.

APPENDIX

For the ground state of a quantum liquid, we may justify (11) and (12) by assuming

(a) the energy of the unperturbed state is given by an expression of the form $N[\epsilon(\rho)/\rho]$, being otherwise independent of the shape and surface area of the boundary enclosing the assembly provided the ratio of volume to surface is large enough compared with a characteristic length $l(\rho)$;[11]

(b) the internal energy of the ground state when it is perturbed by an external field (such as gravity) tends to that of the unperturbed state as the gradient of the field tends to zero; and

(c) the energy of the perturbed ground state is not appreciably affected when the volume is bisected by a thin solid boundary wall, provided the two subvolumes are each large compared with l times the area of the new surface.

Then, choosing k small enough, by (c) we may, without appreciably changing the energy of the assembly perturbed by the potential $\Omega\phi$, partition the volume by thin walls into cells each with linear dimensions small compared with $1/k$. After this move, the perturbed system is replaced by a large number of subsystems, each with a fraction of the original N atoms, with the same total internal energy. On removing the perturbing potential, by (b) the internal energies of the subsystems are not appreciably changed. Hence by (a) the internal energy of the original perturbed system is given by a sum over the cells which may be replaced by the integral $\int\epsilon(\rho)d^3\mathbf{r}$ taken over the whole volume of the assembly; from which (11) follows.

[10] See Whittaker and Watson, *A Course of Modern Analysis* (Cambridge University Press, London, England and Macmillan Publishing Company, New York, 1947), p. 170.

[11] For liquid helium (He^4), l may be as large as $\sim 10^{-4}$ cm (see reference 8).

Atomic Theory of the Two-Fluid Model of Liquid Helium

R. P. FEYNMAN
California Institute of Technology, Pasadena, California
(Received January 11, 1954)

It is argued that the wave function representing an excitation in liquid helium should be nearly of the form $\Sigma_i f(\mathbf{r}_i)\phi$, where ϕ is the ground-state wave function, $f(\mathbf{r})$ is some function of position, and the sum is taken over each atom i. In the variational principle this trial function minimizes the energy if $f(\mathbf{r})=\exp(i\mathbf{k}\cdot\mathbf{r})$, the energy value being $E(k)=\hbar^2k^2/2mS(k)$, where $S(k)$ is the structure factor of the liquid for neutron scattering. For small k, E rises linearly (phonons). For larger k, $S(k)$ has a maximum which makes a ring in the diffraction pattern and a minimum in the $E(k)$ *vs* k curve. Near the minimum, $E(k)$ behaves as $\Delta+\hbar^2(k-k_0)^2/2\mu$, which form Landau found agrees with the data on specific heat. The theoretical value of Δ is twice too high, however, indicating need of a better trial function.

Excitations near the minimum are shown to behave in all essential ways like the rotons postulated by Landau. The thermodynamic and hydrodynamic equations of the two-fluid model are discussed from this view. The view is not adequate to deal with the details of the λ transition and with problems of critical flow velocity.

In a dilute solution of He^3 atoms in He^4, the He^3 should move essentially as free particles but of higher effective mass. This mass is calculated, in an appendix, to be about six atomic mass units.

IN a previous paper,[1] II, a physical argument was given to interpret the fact that the excitations which constitute the normal fluid in the two-fluid theory of liquid helium were of two kinds. Those of lowest energy are longitudinal phonons. The main result of that paper was to give the physical reason for the fact that there can be no other excitations of low energy. It was shown that any others must have at least a minimum energy Δ. No quantitative argument was given to obtain this Δ nor to get an idea of the type of motion that such an excitation represents. In this paper we expect to determine Δ and the character of the excitations.

The physical arguments of II are carried a step further here to show that the wave function must be of a certain form. The form contains a function whose exact character is difficult to establish by intuitive arguments. However, the function can be determined, instead, from the variational principle as that function which minimizes the energy integral.

THE WAVE FUNCTION FOR EXCITED STATES

In II the exact character of the lowest excitation was not determined, but various possibilities were suggested. One is the rotation of a small ring of atoms. A second is the excitation of an atom in the local cage formed around it by its neighbors. Still a third is analogous to the motion of a single atom, with wave number k about $2\pi/a$, where a is the atomic spacing, the other atoms moving about to get out of the way in front and to close in behind. It is not clear that they are really distinct possibilities, for they might be merely different ways of describing roughly the same thing.

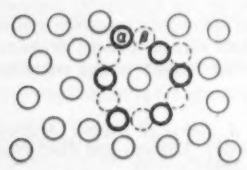

FIG. 1. Typical configuration of the atoms. If an excitation represents rotation of a ring of atoms such as the six in heavy outline the wave function must be plus if they are in the α positions and minus if they are moved to the intermediate β positions.

We shall now try to find the form of the wave function which we would expect under the assumption that one or another of these possibilities is correct. It will turn out that all of the alternatives suggest the same wave function, at least to within a function $f(\mathbf{r})$, of position $\mathbf{r}$, which is determined only vaguely.

First, suppose that the excitation is the rotation of a small ring of atoms. The number of atoms in the ring is determined, according to II, by the condition that it is the smallest ring that can be considered to be able to turn easily as an independent unit in view of the interatomic forces. For illustrative purposes we suppose this means that there are six atoms in the ring.

We can describe the wave function for this excitation by giving the amplitude associated with every configuration of the atoms. Suppose Fig. 1 represents a typical configuration, the six atoms of the ring in question (say ring A) being indicated by heavy outline. We discuss how the amplitude changes as we rotate this ring, leaving the other atoms out of account for a moment. Suppose the wave function is positive, say $+1$, if the atoms are in the position shown by the full circles in Fig. 1, which we arbitrarily call the α position. Suppose all the six atoms move around together, and let the ring turn about 60°. The atoms then appear again in α position, although which is which has been changed, so the wave function, by the Bose statistics, is still $+1$. On the other hand, for a 30° rotation, if the atoms are located as indicated in the figure by dotted circles (β position), the wave function will change to -1 for the first excited state. We need only discuss the real part of the wave function—the imaginary part, if any, can be dealt with in a similar way. (Actually since we deal with an eigenstate of the energy, the real part of

[1] R. P. Feynman, Phys. Rev. 91, 1291, 1301 (1953), hereafter called I, II, respectively.

the wave function is an eigenfunction also.) For orientations intermediate between α, β the function is correspondingly intermediate between $+1$ and -1, but to simplify the remarks we describe it for just the configurations α, β. The wave function for excitation of this ring we call ψ_A. It is $+1$ if the A ring is at α, and -1 if at β, and does not depend on how other rings of atoms are oriented. We can describe this wave function as follows. Consider a function of position $\mathbf{r}$ in space, $f_A(\mathbf{r})$ which is $+1/6$ if $\mathbf{r}$ is at one of the six positions of the centers of the atoms for the α position of ring A, is $-1/6$ if it is at a β position, and is zero if $\mathbf{r}$ is at any other place in the liquid far from the A ring. Then consider the quantity $\sum_i f_A(\mathbf{r}_i)$ where the sum is taken over all the atoms, i, in the liquid. For a configuration of the liquid for which there are atoms at the six α positions the quantity is $+1$, while if six atoms are at β position, it is -1. This suggests that we can write $\psi_A=\sum_i f_A(\mathbf{r}_i)$.

Actually this is incomplete because it does not correctly describe what happens if atoms in other parts of the liquid move. If ring A is in the α position, we wish the complete wave function to be $+1$ as far as this is concerned, but to drop to zero if two atoms overlap in other parts of the liquid, etc., just as for the ground state. That is, we expect (disregarding normalization)

$$\psi_A=\sum_i f_A(\mathbf{r}_i)\phi, \tag{1}$$

where ϕ is the ground-state wave function, a function of all the coordinates. This takes care of another matter also. What happens if some atoms are on α and some on β? This should be of very small amplitude because we do not wish the atoms to overlap on account of the repulsions. This is not correctly described by $\sum_i f_A(\mathbf{r}_i)$, but the ϕ factor does guarantee such a behavior. It is small for such overlaps. Of course, if the ring contained many atoms it could readjust just a little and the ϕ would not prevent, for example, all those near one side of the ring being α, and those on the opposite side of the ring being β. We are not guaranteed that (1) will describe well the amplitude for such a configuration. In fact, it wouldn't be expected that a function of just one variable could describe the motion of several atoms. However, by the arguments of II the ring is supposed to be small, in fact, so small that one part of the ring cannot move independently of the rest. The ring is so small that if one atom is at α, there cannot be a large amplitude for finding atoms at β because of the interatomic repulsions. This is represented in (1) by the factor ϕ which falls if two atoms approach (see II for a full description of the properties of ϕ).

Not knowing the exact size and shape of the ring we cannot say what the exact function $f_A(\mathbf{r})$ should be. But at least we conclude in this case the excited-state wave function is of the form

$$\psi=\sum_i f(\mathbf{r}_i)\phi, \tag{2}$$

where $f(\mathbf{r})$ is *some* function of position.

We might try to improve (1) by noting that, of course, the energy should be essentially the same if the excited ring were somewhere else in the liquid, say at B. The function

$$\psi_B=\sum_i f_B(\mathbf{r}_i)\phi \tag{3}$$

would describe this if $f_B(\mathbf{r})$ is $+1/6$ for $\mathbf{r}$ at some one of the six α positions of some other ring B, and $-1/6$ for intermediate β positions, and zero elsewhere. Or we could locate the ring at still another position, etc. Any one atom might be thought of as belonging to more than one ring. This produces a kind of interaction between adjacent rings. Because of this interaction, a better wave function than (1) might be some linear combination of these possibilities, say $c_A\psi_A+c_B\psi_B+\cdots$. But we can still conclude that the form of the wave function is given by (2), but now, with the function $f(\mathbf{r})=c_A f_A(\mathbf{r})+c_B f_B(\mathbf{r})+\cdots$, for any linear combination of functions of the form (2) is still of this form.

If the lowest excited state which we seek were something like the excitation of a single atom in a cage formed from its neighbors we would guess the wave function to be of the form (2) also. Because there would be a nodal plane across the cage, and we would take $f(\mathbf{r})$ to be positive if $\mathbf{r}$ is in the cage on one side of the plane, and negative if on the other, and to fall off to zero if $\mathbf{r}$ goes outside the cage. We do not care which atom is in the cage so the sum on i is taken over all atoms. Those which are outside the cage contribute nothing to the sum, because $f(\mathbf{r})$ is zero there. Further, there is no appreciable amplitude for there being more than one atom in the cage, because of the action of the factor ϕ which is very small if the atoms penetrate each other's mutual potential. The ϕ also takes care of the fact that the atoms in remote parts of the liquid behave independently of what the excited atom is doing, and act just as in the ground state. Further, linear combinations, representing the alternatives that the excited cage may be located at different places in the liquid, are still of the form (2).

The third possibility was only crudely described in II. It was noted that if the atoms were considered as roughly confined to cells, then a wave function representing the motion of an atom A could be $\exp(i\mathbf{k}\cdot\mathbf{r}_A)$, where $\mathbf{r}_A$ is the position of A, and it is assumed that as A moves about, the other atoms move around to make way for it so that the density is maintained roughly uniform. This would correspond in the liquid to a wave function

$$\exp(i\mathbf{k}\cdot\mathbf{r}_A)\phi, \tag{4}$$

where ϕ is the ground-state wave functions of all the atoms including A. The factor ϕ does the equivalent of keeping the atoms in cells so that the density is nearly uniform no matter where $\mathbf{r}_A$ is. For small $\mathbf{k}$ this is a possibility only if atom A is different from the others and does not obey the Bose statistics. If the symmetry is taken into account then we must replace this by

the symmetrical sum

$$\sum_i \exp(i\mathbf{k}\cdot\mathbf{r}_i)\phi. \tag{5}$$

If ϕ had no large scale density fluctuations this would be no wave function at all, because there would be just as many atoms in the region where $\exp(i\mathbf{k}\cdot\mathbf{r})$ is positive as where it is negative and the sum cancels out.[2] This is in concert with the idea that the wave function cannot depend on where atom A is on a large scale. For if A moves a long distance and the others readjust to keep the density uniform, on a large scale (the scale $1/k$ for small k), the result is just equivalent to the interchange of atoms and the wave function cannot change as a consequence of the Bose symmetry. On the other hand, if while the atom moves from one position to that of its neighbor the wave function changes sign and returns, then (5) may be allowed. That is, something like (5) with k of order $2\pi/a$ may be a possibility. This again is of the form (2), but with $f(\mathbf{r})=\exp(i\mathbf{k}\cdot\mathbf{r})$. The argument just given for this alternative is admittedly not as complete as for the others, mainly because the original idea of what the state is, was based on such a crude model of atoms in cells. Insofar as the idea can be carried over to the case of the true liquid perhaps we can say the form (5), or (2) will represent it.

Since all the examples have led to the same form, we might expect that a more general argument could be made for the validity of (2). This is, in fact, possible starting from the general argument given in II to show why the excited states, other than phonons, can be expected to have an excitation. It was pointed out there that the excited-state function ψ must be orthogonal to the ground state. For some configuration, say α, of the atoms it acquires its maximum positive value. Then it will be negative for some other, say β, which represents some stirring from the α configuration without change of large scale density (to avoid phonon states). But stirring reproduces a configuration nearly like α although with some atoms interchanged. Thus it is hard to get the configuration β to be very far (in configuration space) from α to keep the gradient of ψ small in going from α to β.

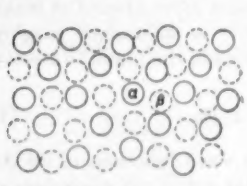

FIG. 2. In general the lowest excitation energy results if the configuration of atoms (solid circles) for which the wave function is most positive is as far as possible from that (dotted circles, β) for which it is most negative. All the β positions must be as far as possible from α positions, therefore.

[2] We shall see later that (5), for small $\mathbf{k}$, is actually a satisfactory wave function because ϕ does have the long wave density variations of the zero point motion of the sound field. We are trying to get excited states orthogonal to phonon states, and (5) for small $\mathbf{k}$ is not orthogonal. It is, in fact, just the wave function for such a phonon state. This is discussed later.

The lowest state would have the β configuration as far as possible from α. This means that in β as many atoms as possible are moved from sites (call them α positions) occupied by atoms in α. Hence β must be a configuration in which the atoms occupy sites (β positions) which are placed as well as possible between the α positions. (See Fig. 2.) In all these configurations, of course, the gross density must be kept uniform and the atoms should be kept from overlapping, to avoid high potential energy terms. If all atoms are on α positions ψ is maximum positive, and if all on β, maximum negative. The transition is made as smoothly as possible, and the kinetic energy thereby kept down, if for other configurations the amplitude is taken to be just the number of atoms on α positions minus the number on β positions. The number is just $\sum_i f(\mathbf{r}_i)$ where $f(\mathbf{r})$ is a function which is $+1$ if $\mathbf{r}$ is at an α position, and -1 if at a β position (and varies smoothly in between these limits as $\mathbf{r}$ moves about). It is of course a modulation to be taken on ϕ, because we wish to give small amplitude to configurations in which atoms overlap, etc., just as in the ground state. We are led, therefore, to (2). We can add the information that $f(\mathbf{r})$ must vary rapidly from plus to minus in distances of half an atomic spacing. That is, we expect that $f(\mathbf{r})$ will consist predominently of Fourier components of wave number $\mathbf{k}$ of absolute magnitude $k=2\pi/a$.

In the above argument it is not self-evident that in going from the configuration of all atoms at α positions to that of all at β, the amplitude must be just linear in the number on α, N_α minus the number of β, N_β. Perhaps some other smooth function of this number, like $\sin[\pi(N_\alpha-N_\beta)/2N]$ might be better. However, for the majority of possible configurations N_α and N_β are nearly equal; in fact, for almost all, $(N_\alpha-N_\beta)/N$ is of order $\pm N^{-\frac{1}{2}}$. For such a small range of the variable, the function, whatever it is, ought to behave nearly linearly. If the wave function (2) is wrong for a very few special configurations it will not be important as we shall determine the energies by the variational method, and the special configurations will contribute only a small amount to the integrals because of their small share of the volume in configuration space.

THE EXCITATION ENERGY

We have concluded that a function of the form (2) should be a good approximation to the wave function of the excited state.[2a] The function $f(\mathbf{r})$ is known only imperfectly, however. We shall determine this function $f(\mathbf{r})$ by using the variational principle. The Hamiltonian of the system is

$$H=-(\hbar^2/2m)\sum_i\nabla^2_i+V-E_0, \tag{6}$$

[2a] Wave functions of this form have been proposed before, for example by A. Bijl, Physica 7, 869 (1940). However, an argument establishing their validity for large $\mathbf{k}$ has been lacking, and it has not been clear that functions of other forms might not give much lower states.

where V is the potential energy of the system, and we measure energies above the ground-state energy E_0, so E_0 is subtracted in (6). Therefore the ground-state wave function satisfies

$$H\phi=0. \tag{7}$$

If we write

$$\psi=F\phi, \tag{8}$$

where F is a function of all the coordinates, then we can verify, using (7), that

$$H\psi=H(F\phi)=-(\hbar^2/2m)\textstyle\sum_i(\phi\nabla_i^2F +2\nabla_i\phi\cdot\nabla_iF)=\phi^{-1}(-\hbar^2/2m)\sum_i\nabla_i\cdot(\rho_N\nabla_iF), \tag{9}$$

where $\rho_N(\mathbf{r}^N)=\phi^2$ is the density function for the ground state, that is, the probability of finding the configuration $\mathbf{r}^N$ (we use $\mathbf{r}^N$ to denote the set of coordinates $\mathbf{r}_i$ of all the atoms, and $\int\cdots d^N\mathbf{r}$ to represent the integral over all of them).

The energy values come from minimizing the integral, (note ϕ is real)

$$\mathcal{E}=\int\psi^*H\psi d^N\mathbf{r} =(\hbar^2/2m)\sum_i\int(\nabla_iF^*)\cdot(\nabla_iF)\rho_Nd^N\mathbf{r}, \tag{10}$$

subject to the condition that the normalization integral,

$$\mathcal{I}=\int\psi^*\psi d^N\mathbf{r}=\int F^*F\rho_Nd^N\mathbf{r}, \tag{11}$$

is fixed. The energy is then $E=\mathcal{E}/\mathcal{I}$.

In these expressions we must substitute

$$F=\textstyle\sum_if(\mathbf{r}_i). \tag{12}$$

Consider the normalization integral first. It is

$$\mathcal{I}=\textstyle\sum_i\sum_j\int f^*(\mathbf{r}_j)f(\mathbf{r}_i)\rho_Nd^N\mathbf{r}.$$

For a fixed i and j we can integrate first over all of the other atomic coordinates. This integral on ρ_N gives the probability for finding the ith atom at $\mathbf{r}_i$ and the jth at $\mathbf{r}_j$; therefore

$$\mathcal{I}=\int f^*(\mathbf{r}_1)f(\mathbf{r}_2)\rho_2(\mathbf{r}_1,\mathbf{r}_2)d^3\mathbf{r}_1d^3\mathbf{r}_2, \tag{13}$$

where ρ_2 is the probability of finding an atom at $\mathbf{r}_1$ per cm^3, and at $\mathbf{r}_2$ per cm^3. These density functions can be defined in general by

$$\rho_k(\mathbf{r}_1',\mathbf{r}_2'\cdots\mathbf{r}_k')=\textstyle\sum_i\sum_j\cdots\sum_n\int\delta(\mathbf{r}_i-\mathbf{r}_1') \times\delta(\mathbf{r}_j-\mathbf{r}_2')\cdots\delta(\mathbf{r}_n-\mathbf{r}_k')\rho_N(\mathbf{r}^N)d^N\mathbf{r}. \tag{14}$$

For example, $\rho_1(\mathbf{r})$ is simply the chance of finding an atom at $\mathbf{r}_1'$, for the liquid in the ground state. This is independent of $\mathbf{r}$ and is the number density ρ_0 in the ground state. In the same way $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$ can be written as $\rho_0p(\mathbf{r}_1-\mathbf{r}_2)$ where p is the probability of finding an atom at $\mathbf{r}_2$ per unit volume if one is known to be at $\mathbf{r}_1$. Except near the liquid surface it is a function of only the distance from $\mathbf{r}_1$ to $\mathbf{r}_2$, so (13) is

$$\mathcal{I}=\rho_0\int f^*(\mathbf{r}_1)f(\mathbf{r}_2)p(\mathbf{r}_1-\mathbf{r}_2)d^3\mathbf{r}_1d^3\mathbf{r}_2. \tag{15}$$

The energy integral (10), with the substitution (12) becomes

$$\mathcal{E}=(\hbar^2/2m)\textstyle\sum_i\int\nabla_if^*(\mathbf{r}_i)\cdot\nabla_if(\mathbf{r}_i)\rho_Nd^N\mathbf{r}.$$

The integral of ρ_N over all atomic coordinates except $\mathbf{r}_i$ gives a result involving only $\rho_1(\mathbf{r}_i)=\rho_0$. Therefore we have simply

$$\mathcal{E}=\rho_0(\hbar^2/2m)\int\nabla f^*(\mathbf{r})\cdot\nabla f(\mathbf{r})d^3\mathbf{r}. \tag{16}$$

The best choice of f is that which minimizes the ratio of (16) to (15). The variation with respect to f^* gives the equation

$$E\int p(\mathbf{r}_1-\mathbf{r}_2)f(\mathbf{r}_2)d^3\mathbf{r}_2=-(\hbar^2/2m)\nabla^2f(\mathbf{r}_1),$$

where the energy E is $\mathcal{E}/\mathcal{I}$. This has the solution

$$f(\mathbf{r})=\exp i(\mathbf{k}\cdot\mathbf{r}), \tag{17}$$

with the energy value

$$E(\mathbf{k})=\hbar^2k^2/2mS(\mathbf{k}), \tag{18}$$

where $S(\mathbf{k})$ is the Fourier transform of the correlation function,

$$S(\mathbf{k})=\int p(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r})d^3\mathbf{r}. \tag{19}$$

It is a function only of k, the magnitude of $\mathbf{k}$.

It is readily verified that the solution is orthogonal to the ground state if we exclude $\mathbf{k}=0$. In fact, the solutions for different values of $\mathbf{k}$ are orthogonal to each other. This is because they all belong to different eigenvalues, $\hbar\mathbf{k}$, of the total momentum operator

$$\mathbf{P}=(\hbar/i)\textstyle\sum_i\nabla_i,$$

as is directly verified from (2) with (17), taking $\mathbf{P}\phi=0$ since the ground state has zero total momentum.[3] Since

[3] The argument is not rigorous because the momentum of the entire liquid can be changed without appreciable energy change by moving the center of gravity. This multiplies the wave function by a factor like $\exp(-i\hbar\mathbf{k}N^{-1}\cdot\Sigma_i\mathbf{r}_i)$. This function is so different from (2), however, that the orthogonality is probably not destroyed.

this operator commutes with the Hamiltonian, we have in (18) an upper limit to the energy for each value of k. (In fact, we could have obtained (17) from (2) by this argument.) Since we expect that (2) is a good wave function for functions f which vary from plus to minus in a distance of order $a/2$, we expect that (18) is not only an upper limit, but also a good estimate of the energy in a range of k's in the neighborhood of $k=2\pi/a$. In fact, our arguments suggest that $E(\mathbf{k})$ should have a minimum as a function of k in that region. These expectations are verified in the next section.

DISCUSSION OF THE ENERGY SPECTRUM

To find the consequences of (18) we shall have to discuss the behavior of $S(k)$ defined in (19).

The function $p(\mathbf{r}_1-\mathbf{r}_2)$ gives the probability per unit volume that a particle is at $\mathbf{r}_2$ if one is known to be at $\mathbf{r}_1$. If $\mathbf{r}_2$ is close to $\mathbf{r}_1$ it is zero, for the atoms cannot overlap. On the other hand, if $\mathbf{r}_2$ coincides with $\mathbf{r}_1$ there is an atom there, so p contains a delta function $\delta(\mathbf{r}_1-\mathbf{r}_2)$. For large r_2 it approaches ρ_0. Since the structure of the liquid ought to be more or less like that in a classical fluid, as r increases from zero, $p(r)$ probably rises to a maximum at the nearest neighbor spacing, falls, then rises again to a lower and wider maximum for next nearest, and with rapidly decreasing smaller oscillations approaches unity.[4] The integral $\int(p(\mathbf{r})-1)d^3\mathbf{r}$ vanishes since the integral of $\rho_2(\mathbf{r}_1,\mathbf{r}_2)$ with respect to $\mathbf{r}_2$ is exactly ρ_1 times the number of atoms $N=\rho_0 V$.

The Fourier transform function $S(k)$ is just the liquid structure factor which determines the scattering of neutrons (or x-rays, after multiplication by the atomic structure factor) by the liquid at absolute zero. It is therefore a quantity which can be directly determined experimentally. For large k it approaches 1 because of the delta function in $p(\mathbf{r})$. It has a delta function at $\mathbf{k}=0$, but this value of k is not of interest to us in (18), because the wave function ψ must be orthogonal to the ground state. The behavior at small k depends on the variations of $p(r)$ over long distances, that is, on long wavelength density fluctuations. These are the zero point fluctuations of the sound field in the ground state, since for wavelengths longer than the atomic spacing the approximation of a continuous sound field is good. This may be analyzed as follows. The operator representing the density at a point $\mathbf{r}$ is

$$\rho(\mathbf{r})=\sum_i\delta(\mathbf{r}-\mathbf{r}_i). \tag{20}$$

Its Fourier transform is

$$q_{\mathbf{k}}=\int\rho(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r})d^3\mathbf{r}=\sum_i\exp(i\mathbf{k}\cdot\mathbf{r}_i). \tag{21}$$

Evidently, the $S(\mathbf{k})$ is the expected value of $|q_{\mathbf{k}}|^2$ in the ground state. For long wave sound $q_{\mathbf{k}}$ is just the coordinate of the normal mode, so its mean square can be easily determined, for example, by noting that the mean potential energy is half of the ground-state energy $\frac{1}{2}\hbar\omega$. In this way one finds $S(\mathbf{k})=\hbar k/2mc$ for small k, where c is the velocity of sound.

The behavior of $S(\mathbf{k})$ for intermediate k is familiar to us from the x-ray studies of classical liquids. The density distribution in the ground state is roughly similar to such a liquid. There is some local structure produced by the tendency of the atoms to stay apart. This quasi-crystalline local order makes a maximum in the $S(k)$ curve for k near $2\pi/a$. There may be smaller subsidiary maxima for near multiples of this k. For helium, because of the large zero point motion, these maxima may be broader and less marked than in other liquids. The main maximum is responsible for the main ring in the x-ray diffraction pattern. It is shown clearly in the preliminary neutron diffraction data reported by Henshaw and Hurst.[5]

To summarize: with rising k, $S(k)$ starts linearly as $\hbar k/2mc$, rises then to a maximum near $k=2\pi/a$, and falls again to approach, with possible minor oscillations, the limit unity. Consequently the quantity $E(k)=\hbar^2k^2/2mS(k)$ should start linearly as $\hbar kc$, but should then show a dip with a minimum at $k=k_0$ say, near $2\pi/a$, finally rising, eventually as $\hbar k^2/2m$. These relations are shown in Fig. 3.

We have argued that (2) should be a good approximation to the wave function for functions that contain wave numbers in the vicinity of $2\pi/a$. Therefore we can expect the energy values (18) to be good in the neighborhood of this wave number. It is gratifying to see that there is a minimum in this region. The minimum value we shall call Δ. Ordinarily the variational method only permits one to interpret the minimum value of E as one varies a parameter such as k. On the other hand, in our case each value of k has significance since these values correspond to different eigenvalues of the momentum operator, as has been remarked. Therefore we can believe the behavior of the curve through a range of k near k_0, where it behaves parabolically, so we can write $E(k)$ in Landau's form $\Delta+\hbar^2(k-k_0)^2/2\mu$ where μ is a constant determining the curvature.

It is at first disconcerting that values of the energy lower than Δ can be obtained by going to very small values of k. But the energy here varies as $\hbar kc$, just that expected for phonon excitation. In fact, a moment's reflection shows that, for small k, the wave function (2) is just that which represents phonon excitation. Excitation of a given phonon means that the harmonic oscillator representing the corresponding normal mode is in the first excited state. The wave function is therefore $q_{\mathbf{k}}\phi$, if $q_{\mathbf{k}}$ is the normal coordinate of the mode excited. This coordinate is the Fourier transform of the density, so (21) shows that (2) with (17) represents a phonon for small k. Since the wave function is correct the energy must be exact, and is therefore $\hbar kc$.

[4] J. Reekie and T. S. Hutchison, Phys. Rev. **92**, 827 (1953), have determined $p(r)$ by x-ray scattering.

[5] D. G. Henshaw and D.G. Hurst, Phys. Rev. **91**, 1222 (1953).

Although we have made an argument only to show that (2) should be valid for high k, we see now that it is also valid for small k, that is for $f(\mathbf{r})$ which vary slowly. Since the energy curve is valid for the smaller k and for a range about $2\pi/a$, we can accept it as reasonable for all k from zero up to and slightly beyond the minimum.

On the other hand, for still larger k, another state of lower energy exists with the same total momentum. It is the state of double excitation, one of $\mathbf{k}_1$, the other of $\mathbf{k}_2$, such that $\mathbf{k}_1+\mathbf{k}_2=\mathbf{k}$ and still $E(\mathbf{k}_1)+E(\mathbf{k}_2)<E(\mathbf{k})$. This becomes possible for k so high that the slope dE/dk of the energy curve exceeds $\hbar c$, the initial slope. The curve for very large k, therefore, does not have the same validity as that for lower k, but we need not enter into this matter, because at temperatures of a few degrees such high-energy states would not be appreciably excited. Such questions may be of importance in discussing nonequilibrium phenomena. One process by which the number of excitations can change is for an excitation to pick up enough momentum that it can divide spontaneously into two.

It is easy to misinterpret the meaning of the wave function

$$\psi=\sum_i \exp(i\mathbf{k}\cdot\mathbf{r}_i)\phi, \tag{22}$$

so a few remarks might be appropriate here. It looks at first, on inspection of the first factor, that this represents the excitation of a single particle. This is correct at very high $k(ka\gg 2\pi)$ and it is also correct for the ideal gas case for which the atoms do not interact (ϕ is constant then). But our arguments for intermediate k show that this is not the case. Because of the correlations in position implied by the factor ϕ, the motion of one atom implies the motion of others. Thus the factor in front of ϕ selects from that function certain correlated motions, in spite of the fact that each term in the factor depends on just one variable.

We can get a better idea of how this works by taking the extreme case of very low k. Here (22) represents a sound wave but at first sight there is no sign of the density variations that such a wave usually brings to mind. Let us take the real part and consider

$$\sum_i \cos(\mathbf{k}\cdot\mathbf{r}_i)\phi \tag{23}$$

for small k. Now, for most configurations, allowed by ϕ, the atoms are fairly uniformly distributed, so that there are just as many in the regions where the cosine is positive, as where it is negative. Therefore the sum over all the atoms of $\cos(\mathbf{k}\cdot\mathbf{r}_i)$ is zero. The wave function is zero for nearly all configurations. It is only for the rare configurations in which the number in positive regions exceeds that in regions where the cosine is negative that the wave function does not vanish. In this way (23) selects configurations for which the mean density varies as $\cos(\mathbf{k}\cdot\mathbf{r})$. Since such density fluctuations are, according to the behavior of ϕ, most likely produced by small cooperative motions of large numbers of atoms, the state described is very far from the one particle state it would be if the cosine factor appeared alone, not multiplied by ϕ.

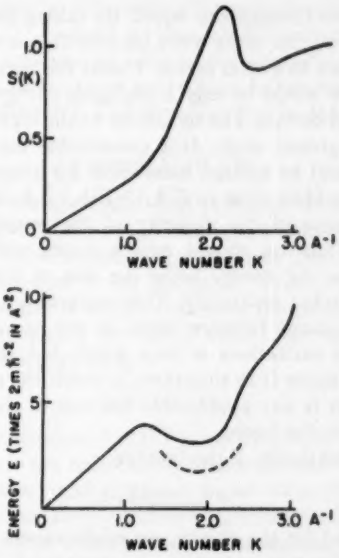

FIG. 3. The upper curve gives the liquid structure factor determined from neutron diffraction (reference 5) and extrapolated to zero k. The lower curve gives the energy spectrum of excitations as a function of wave number (momentum $\cdot\hbar^{-1}$) which results from the formula $E=\hbar^2k^2/2mS(k)$ derived in the text. The initial linear portion represents excitation of phonons while excitations near the minimum of the curve, where it behaves as $\Delta+\hbar^2(k-k_0)^2/2\mu$, correspond to Landau's rotons. However, data on the specific heat indicate that the theoretical curve should lie lower, closer to the dashed curve.

In the region of the energy minimum at k_0 the wave function represents a situation intermediate between the cooperative motion of phonons, and the excitation of a single particle. Several atoms move together because of the correlations implied by ϕ. It is hard to make a clear picture out of this vague idea. There is nothing to indicate that the state carries an intrinsic angular momentum. One must be careful because the state is degenerate, as all directions of k with the same magnitude k_0 give the same energy Δ. Perhaps, if more complicated wave functions were tried, some special linear combination representing a kind of microscopic vortex ring or one with intrinsic angular momentum has in fact a lower energy. States of low k will be called phonons, and states of momentum near k_0 will be called rotons in this paper, in accordance with the terminology of Landau,[6] although we do not necessarily mean to imply that rotons carry intrinsic angular momentum or represent vortex motion.

MULTIPLE EXCITATION

We have obtained the energy spectrum $E(k)$ of what we may call single excitations. They have the form of

[6] L. Landau, J. Phys. U.S.S.R. **5**, 71 (1941); **8**, 1 (1941). See also R. B. Dingle, Supplement to Phil. Mag. **1**, 112 (1952).

plane waves through the liquid. By taking linear combinations we can make wave packets that are more or less confined to a local region. Unless the region is very small there would be only a negligible energy addition required to do this. The remainder of the liquid is quiet as in the ground state. It is conceivable that another packet could be located somewhere far away and the energy would be close to $E(\mathbf{k}_1)+E(\mathbf{k}_2)$ if $\mathbf{k}_1$ and $\mathbf{k}_2$ are the momenta of the majority of the waves in each packet. Thus we should expect states with several excitations, the energy being the sum of the energies of each packet separately. This neglects a kind of interaction energy between them. It will be valid if the density of excitations is very small, but one cannot expect to apply it to situations in which the number of excitations is any appreciable fraction of the number of atoms in the liquid.

Mathematically, if the function

$$F=\sum_i \exp i\mathbf{k}_1\cdot\mathbf{r}_i$$

gives $E(k_1)$ for the energy, we might expect the wave function for two excitations to correspond to $\psi=F\phi$, with

$$F=[\textstyle\sum_i \exp(i\mathbf{k}_1\cdot\mathbf{r}_i)][\sum_j \exp(i\mathbf{k}_2\cdot\mathbf{r}_j)]. \quad (24)$$

It is readily verified, by substitution into the variation integral, that the energy is $E(\mathbf{k}_1)+E(\mathbf{k}_2)$ within correction terms of order $1/V$, where V is the volume of the entire fluid. This is just what one would expect if the excitations behaved like interacting particles, for the relative probability of their being within their range of interaction varies inversely as the volume. The expression (24) is unaltered on reversing the order of the factors, so the state in which the first excitation has momentum $\mathbf{k}_1$ and the second has $\mathbf{k}_2$ is the same as that in which the momenta are reversed. Thus the excitations obey Bose statistics. The expression (24) is not orthogonal to (17) with $\mathbf{k}=\mathbf{k}_1+\mathbf{k}_2$, so there undoubtedly are matrix elements between states of different numbers of excitations, and collisions must be possible which change this number. In summary, the excitations behave much like interacting Bose particles which may be created and destroyed, and whose energy as a function of momentum is given by $E(k)=\hbar^2k^2/2mS(k)$.

THERMODYNAMIC PROPERTIES OF HELIUM II

From this we may determine the thermodynamic behavior of liquid helium at low temperature. At sufficiently low temperatures the number of excitations will be small, so the interactions between them can be neglected. The approximation of independence leads in the usual way to the formula for the Gibbs free energy (taking the ground-state energy as zero),

$$F=kTV\int \ln[1-\exp(-\beta E[\mathbf{k}])]d^3\mathbf{k}(2\pi)^{-3}, \quad (25)$$

with $\beta=1/kT$. The number of excitations of momentum $\mathbf{k}$ is

$$n_\mathbf{k}=[\exp\beta E(\mathbf{k})-1]^{-1}. \quad (26)$$

We need not enter into further details as this has been thoroughly analyzed by Landau,[6] who first proposed the form of energy spectrum we have deduced here. At low temperatures only the lowest energy excitations can become excited. That is, only the phonons are excited and the specific heat varies as T^3. At higher temperatures some of the states near the minimum of the curve, at k_0 become excited. The specific heat then rises rapidly, controlled predominantly by the $\exp(-\beta\Delta)$ factor, governing the number of rotons excited. For temperatures of a few degrees few rotons are excited and only the phonon part, and the part of the curve near the minimum, are important. Landau[7] has shown that one obtains good agreement with the specific heat (and with the measured values of the velocity of second sound) if one chooses the parameters $\Delta=9.6°\text{K}$, $k_0=1.95\ \text{A}^{-1}$, and $\mu=0.77$. This means the energy curve near the minimum behaves as $2mE/\hbar^2=1.6+1.3(k-1.95)^2$, with k in reciprocal angstroms (one A^{-2} corresponds to a temperature of 6°K). In the phonon region the curve is

$$2mE/\hbar^2=2.6k,$$

in the same units, if the speed of sound is 240 meters/sec. Henshaw and Hurst[8] have published some preliminary data on the neutron scattering by liquid helium at 4.2°K. From it $S(k)$ may be directly determined (see Fig. 3). The curve for $E(k)$ calculated in this way behaves as

$$2mE/\hbar^2=3.0+1.0(k-2.0)^2$$

near the minimum (and is consistent with $2.6k$ for small k). This corresponds to a value of Δ of 18° which is impossibly large. Such a discrepancy may be due to the inaccuracy of the trial function (2), the true energy being lower than that calculated with this trial function. Such a large discrepancy in energy is discouraging, because the physical arguments did seem to indicate that (2) should be a reasonably good first approximation.

The expression (25) should not hold at high temperature because it neglects the interactions among the large number of excitations which (26) demands at such a temperature. Without an estimate of these interactions it is hard to judge the region in which deviations are to be expected. We shall make a very rough preliminary argument here.

To the approximation that the energy in a mode is proportional to an integer n, this mode behaves like a harmonic oscillator. The coordinate of this oscillator $q_\mathbf{k}$ has a mean square value $2n+1$ times its value in the ground state. For what size $q_\mathbf{k}$ is the harmonic oscillator approximation poor? If we knew this we could put a limit on the ranges of $q_\mathbf{k}$ and hence of $n_\mathbf{k}$, for which (25) might be expected to be valid. In our case the various $q_\mathbf{k}$ from (21) are not independent, because they can all be defined in terms of the same $3N$ variables $\mathbf{r}_i$.

[7] L. Landau, J. Phys. U.S.S.R. **11**, 91 (1947); Phys. Rev. **75**, 884 (1949).

Thus, for example, if $q_{\mathbf{k}}$ is known for $3N$ values of $\mathbf{k}$, it is known in principle for all others. It is very difficult to see what this interdependence means to (25).

But one can notice that one restriction is

$$\int |q_{\mathbf{k}}|^2 d^3\mathbf{k}(2\pi)^{-3}=N, \tag{27}$$

where the integral is taken over all $\mathbf{k}$. Hence we may guess that we should restrict (25) by the condition that the excess $|q_{\mathbf{k}}|^2$ over that for the ground state, summed on all states, cannot exceed N. This excess for a given mode is $2n_{\mathbf{k}}$ times the ground-state mean value of $|q_{\mathbf{k}}|^2$. This latter is $S(k)=\hbar^2k^2/2mE(k)$, so we obtain the restriction

$$\hbar^2m^{-1}\int (k^2/E(k))n_{\mathbf{k}}d^3\mathbf{k}(2\pi)^{-3}=1. \tag{28}$$

The thermodynamics which would result (by adding a chemical potential μ to $E(k)$ in (25), (26) to allow for (28)) will show a second-order transition.[8] But without a deeper analysis we are in no position to take Eq. (28) literally. We can only use it as a rough criterion for validity of (25). If the integral on the left is much less than 1, then (25) should hold. At 2°K the integral amounts to roughly 0.2, so perhaps we are entitled to trust (25) even to within a few tenths of a degree of the transition temperature.

MOTION OF THE FLUID AS A WHOLE

The existence of such excitations moving as nearly free particles in a background fluid is the central concept of the two-fluid model of Tisza[9] and Landau. The consequences of these ideas for excitations with a spectrum such as (18) have been carefully analyzed in a general manner by Landau and Dingle.[6] There is nothing to add that is new in this direction. However, we shall review briefly how the equations of this model arise, emphasizing the behavior of the wave function.

Beside the states which represent local internal excitation of part of the fluid, there are, of course, states in which the entire body of fluid moves. In general, in these cases the boundaries of the fluid move also. For example, at absolute zero, the entire fluid may move as a body with velocity $\mathbf{v}$. This center-of-gravity motion is described by

$$\psi=\exp(im\mathbf{v}\cdot\textstyle\sum_i\mathbf{r}_i)\phi,$$

if we assume ϕ corresponds to the ground state at rest in the laboratory system.

Suppose we wish to represent a situation in which the velocity $\mathbf{v}(\mathbf{r})$ varies from point to point, but only very gradually on an atomic scale. We might try something like this. The atoms in a region about some point P have their center of gravity moving at velocity $\mathbf{v}_P$ corresponding to this point. They must contribute a phase $m\mathbf{v}_P\cdot\sum_i\mathbf{r}_i$ where the sum is taken only over those near P. Corresponding contributions would come from sums near other points so the total factor ought to be $\exp[im\sum_i\mathbf{v}(\mathbf{r}_i)\cdot\mathbf{r}_i]$. The wave function is, therefore, of the form

$$\psi=\exp[i\textstyle\sum_i s(\mathbf{r}_i)]\phi, \tag{29}$$

where $s(\mathbf{r})$ is some function of position. We have suggested that it is $m\mathbf{v}(\mathbf{r})\cdot\mathbf{r}$. However, as is usual when one has waves whose wavelength varies from point to point, the wave number is not the phase divided by r, but more accurately it is the gradient of the phase. Therefore (29) represents the fluid in motion, the velocity at any point being given by

$$\mathbf{v}(\mathbf{r})=m^{-1}\nabla s. \tag{30}$$

As a consequence of (31), $\nabla\times\mathbf{v}=0$. Velocity fields for which this is not true cannot be represented in such a simple manner, and represent, as we have seen in II, states involving large numbers of excitations. The problem they present is being studied. For regions which are not simply connected, such as a torus, s need not be single-valued. For example, in the torus we could take $s=\phi$, the cylindrical angle. This would represent a permanent circulation[10] even though $\nabla\times\mathbf{v}=0$ locally.

Substitution of (29) into the variational principle to obtain a steady-state solution leads [see (10), (11) with $F=\exp(i\sum_i s(r_i))$] to the energy expression

$$\frac{\rho_0}{2m}\int \nabla s(\mathbf{r})\cdot\nabla s(\mathbf{r})d^3\mathbf{r}, \tag{31}$$

which is the kinetic energy $\rho_0 m\mathbf{v}^2/2$ per unit volume. It is minimum for variations in s if $\nabla\cdot(\nabla s)=0$, that is, $\nabla\cdot\mathbf{v}=0$. The flow must be incompressible. We have not allowed, in (29), for variations in density. For a singly connected region this has but one solution $\mathbf{v}=0$, unless the boundaries move. In a multiply connected region, like a ring, circulation of angular momentum in multiples of $N\hbar$ is possible. There are so few of these special states that the statistical mechanics is not affected. The variables, $\mathbf{v}(\mathbf{r})$, representing such motions can be specified as external known variables like pressure and volume.

EXCITATIONS IN A MOVING FLUID

Next we study the motion and energy of an excitation in a moving fluid. The wave function is

$$\psi=\textstyle\sum_i f(\mathbf{r}_i)\exp[i\sum_j s(\mathbf{r}_j)]\phi.$$

[8] This conclusion is modified if the interaction of the rotons and the hydrodynamic modes is taken into account.

[9] L. Tisza, Phys. Rev. 72, 838 (1947).

[10] It was suggested in II, reference 9, that to observe this experimentally one might have to avoid letting the liquid have a free surface. But R. Peierls has pointed out (private discussion) that although the atoms evaporating from the moving liquid to the gas carry angular momentum out, only those of the gas which are moving along with the liquid can condense, bringing back angular momentum—so in equilibrium there would be no damping of the motion from this effect. There are other effects, however, such as those which cause resistance to capillary flow at high velocities, which might be expected to damp the motion.

When this is put into the variational principle, one finds directly

$$\mathcal{E}/s = E(k) + m\int \mathbf{j}(\mathbf{r})\cdot\mathbf{v}_s(\mathbf{r})d^3\mathbf{r} + \tfrac{1}{2}m\int \rho(\mathbf{r})\mathbf{v}_s(\mathbf{r})\cdot\mathbf{v}_s(\mathbf{r})d^3\mathbf{r}, \quad (32)$$

where $E(k)$ is given in (18), $\mathbf{v}_s(\mathbf{r}) = m^{-1}\nabla s$, and we have defined

$$\rho(\mathbf{a}) = \int \rho_2(\mathbf{a},\mathbf{r}_1,\mathbf{r}_2) f^*(\mathbf{r}_1) f(\mathbf{r}_2) d^3\mathbf{r}_1 d^3\mathbf{r}_2/I, \quad (33)$$

$$\mathbf{j}(\mathbf{a}) = h\int \rho_2(\mathbf{a},\mathbf{r})[f^*(\mathbf{r})\nabla f(\mathbf{a}) - f(\mathbf{r})\nabla f^*(\mathbf{a})]d^3\mathbf{r}/Imi, \quad (34)$$

with

$$I = \int \rho_2(\mathbf{r}_1,\mathbf{r}_2) f^*(\mathbf{r}_1) f(\mathbf{r}_2) d^3\mathbf{r}_1 d^3\mathbf{r}_2. \quad (35)$$

These ρ, $\mathbf{j}$, are the expected values of density and current density that belong to the state representing a single excitation.

We shall analyze this for the case of a single excitation of nearly definite momentum in the form of a large wave packet, large compared with the central wavelength. That is, we take $f(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r})g(\mathbf{r})$ where $g(\mathbf{r})$ is a smooth amplitude function, such as a Gaussian, with width very large compared to $1/k$, but small compared to the size of the vessel. Such a packet will drift and spread slowly in a way completely determined by $E(k)$ and the principle of superposition. We wish to determine the additional effects of the possibility of general liquid flow.

The current density associated with this excitation is found from (34). We make the approximation that $\nabla f = i\mathbf{k}f$ since g varies so slowly, obtaining,

$$\mathbf{j}(\mathbf{a}) = \hbar\mathbf{k}m^{-1}\int \rho_2(\mathbf{a},\mathbf{r})\{g^*(\mathbf{r})g(\mathbf{a})\exp[-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{a})] + g(\mathbf{r})g^*(\mathbf{a})\exp[+i\mathbf{k}\cdot(\mathbf{r}-\mathbf{a})]\}d^3\mathbf{r}/I.$$

Now, because of the variation of the exponentials, contributions to this integral come only from $\mathbf{r}$ within a limited distance from $\mathbf{a}$. Within such a distance $g(\mathbf{r})$ is nearly the same as $g(\mathbf{a})$, so all the g factors can be evaluated at $\mathbf{a}$ and taken outside the integral. The integral on $\mathbf{r}$ is then easy by (19) and one finds

$$\mathbf{j}(\mathbf{a}) = \hbar\mathbf{k}\rho_0 S(k)|g(\mathbf{a})|^2/mI.$$

The normalization integral may be done in a similar manner. It is

$$I = \int\int g^*(\mathbf{r}_1)g(\mathbf{r}_2)\rho_2(\mathbf{r}_1,\mathbf{r}_2)\times\exp[-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)]d^3\mathbf{r}_1 d^3\mathbf{r}_2.$$

Since $\mathbf{r}_2$ must be near $\mathbf{r}_1$ for a large contribution, we may replace $g(\mathbf{r}_2)$ by $g(\mathbf{r}_1)$, integrate $\mathbf{r}_2$ directly and obtain

$$I = \rho_0 S(k)\int |g(\mathbf{r})|^2 d^3\mathbf{r}.$$

If we assume $g(\mathbf{r})$ is normalized, $\int|g(\mathbf{r})|^2d^3\mathbf{r} = 1$, so that $|g(\mathbf{a})|^2 = d(\mathbf{a})$ is the density in the packet at $\mathbf{a}$, or roughly the probability of finding the excitation at $\mathbf{a}$, the current is

$$\mathbf{j}(\mathbf{a}) = (\hbar\mathbf{k}/m)d(\mathbf{a}), \quad (36)$$

that is, a total current $\hbar\mathbf{k}/m$ distributed at density $d(\mathbf{a})$.

The particle density at $\mathbf{a}$ is

$$\rho(\mathbf{a}) = \int \rho_2(\mathbf{a},\mathbf{r}_1,\mathbf{r}_2)g^*(\mathbf{r}_1)g(\mathbf{r}_2)\times\exp[-i\mathbf{k}\cdot(\mathbf{r}_1-\mathbf{r}_2)]d^3\mathbf{r}_1 d^3\mathbf{r}_2/I. \quad (37)$$

The points $\mathbf{r}_1$ and $\mathbf{r}_2$ must be close. If point $\mathbf{a}$ is not close to these points we can use the asymptotic form,

$$\rho_2(\mathbf{a},\mathbf{r}_1,\mathbf{r}_2) = \rho_0\rho_2(\mathbf{r}_1,\mathbf{r}_2), \quad (38)$$

to show directly that $\rho(\mathbf{a})$ is the density ρ_0 of the fluid, far from the packet. It is nearly so, even in the region of the packet, for since its dimensions are large, $\mathbf{a}$ is nearly always far from $\mathbf{r}_1,\mathbf{r}_2$, and further, the integral over all $\mathbf{a}$ of $\rho_2(\mathbf{a},\mathbf{r}_1\mathbf{r}_2) - \rho_0\rho_2(\mathbf{r}_1,\mathbf{r}_2)$ is exactly zero.

It is true that the distance of influence in ρ may not be very small, because of the correlations in the sound field. That is, the excitation produces a small strain in the fluid which makes a field of stress in the vicinity. Such fields provide a mechanism of interaction between excitations (as well as a correction to the energy of one). In a more detailed analysis such effects should be taken into account. Here we proceed to a first approximation and neglect them. To the approximation of neglecting compressibility, then, we find $\rho(\mathbf{a}) = \rho_0$; the presence of an excitation does not change the fluid density.

Thus we picture an excitation in the form of a drifting wave packet as carrying a total current $\hbar\mathbf{k}/m$, and drifting (if $\mathbf{v} = 0$) at the group velocity $\mathbf{v}_g = \partial E/\partial\mathbf{k}$, but as not appreciably altering the density.

This clearly violates the conservation of matter. For a moment we overlook this difficulty. It is discussed in the section following the next.

If this packet is in a general velocity field $\mathbf{v}_s(\mathbf{r})$ we may determine its energy from (32). In integral $\int\mathbf{j}(\mathbf{r})\cdot\mathbf{v}_s(\mathbf{r})d^3\mathbf{r}$ we shall assume that $\mathbf{v}_s$ does not vary appreciably over a region as small as the packet, and may be taken outside the integral sign. The integral of $\mathbf{j}$ is then $\hbar\mathbf{k}/m = \mathbf{p}/m$, giving the following results:

The energy of an excitation in a moving fluid is

$$E = E(p) + \mathbf{p}\cdot\mathbf{v}_s, \quad (39)$$

where $\mathbf{v}_s$ is the velocity of the fluid where the excitation (considered as a packet) is located. The total momentum

associated with the packet is $\mathbf{p}$, and it contributes a current $\mathbf{p}/m$ to the total in the fluid. The energy contributed by the moving fluid has density $\rho_0 m\mathbf{v}_s^2/2$, and it contributes to the current density $\rho_0\mathbf{v}_s$.

The group velocity of the excitation is $\partial E/\partial\mathbf{p}$ so that $\mathbf{v}_g=\mathbf{v}_{g0}+\mathbf{v}_s$ where $\mathbf{v}_{g0}$ is the group velocity in liquid at rest, $\partial E(p)/\partial\mathbf{p}$. Thus the excitation just drifts along with the background fluid motion, of velocity $\mathbf{v}_s$. Equation (39) can be obtained much more simply by a Galilean transformation of coordinates. Indeed, it is in this way that it was obtained by Landau and Dingle.[6]

RELATION TO THE TWO-FLUID MODEL

The two paragraphs at the end of the previous section contain the main relations by which the hydrodynamics and thermodynamics of the two-fluid model is derived. We review here a few of the steps, very briefly, in order to make clear the relation of the excitations to what is called the normal fluid. For further details see reference 6.

Thermodynamic equilibrium results, for a system with Bose statistics, if the excitations are distributed so that the number of those with energy E is

$$n=[\exp(\beta E)-1]^{-1}.$$

This may be obtained, for example, by maximizing the entropy, keeping the total energy constant. Another distribution which is also in equilibrium can be got by maximizing the entropy, keeping both total energy and total momentum constant. It is

$$n=\{\exp[\beta(E-\mathbf{p}\cdot\mathbf{u})]-1\}^{-1}, \tag{40}$$

where $\mathbf{u}$ is a constant.

In our liquid, the density of excitations per unit volume at a point $\mathbf{r}$ in this case would be, substituting (39) into (40),

$$n=\{\exp[\beta(E(\mathbf{p})+\mathbf{p}\cdot\mathbf{v}_s(\mathbf{r})-\mathbf{p}\cdot\mathbf{u})]-1\}^{-1}. \tag{41}$$

In order to interpret $\mathbf{u}$, we study the total current density. Since each excitation contributes a current $\mathbf{p}/m$ the total current density contributed by the excitations is

$$m^{-1}\int\mathbf{p}\{\exp[\beta(E(\mathbf{p})+\mathbf{p}\cdot(\mathbf{v}_s-\mathbf{u}))]-1\}d^3\mathbf{p}(2\pi)^{-3}.$$

At this stage we shall only consider the case of low macroscopic velocities. Expanding to the first order in in $\mathbf{v}_s-\mathbf{u}$ this may be written in the usual way as

$$\rho_n(\mathbf{u}-\mathbf{v}_s),$$

where ρ_n is defined as

$$\rho_n=\frac{\beta}{3m}\int p^2\{\exp[\beta E(p)]-1\}^{-2}d^3\mathbf{p}(2\pi)^{-3}. \tag{42}$$

To this we must add the current of the background $\rho_0\mathbf{v}_s$, so that the total macroscopic current density can be written

$$\mathbf{j}=\rho_n\mathbf{u}+\rho_s\mathbf{v}_s, \tag{43}$$

if we put $\rho_s=\rho_0-\rho_n$.

In view of these separations we can say, artificially, that the liquid behaves as though there were two parts, superfluid at density ρ_s moving at velocity $\mathbf{v}_s$, and normal at density ρ_n and velocity $\mathbf{u}$ (which we write hereafter as $\mathbf{v}_n$). The current is the sum of these two partial currents. In a similar manner the change of the internal energy, at constant entropy, produced by the velocities can be shown to second order to be the sum of the kinetic energies $\frac{1}{2}\rho_s\mathbf{v}_s^2+\frac{1}{2}\rho_n\mathbf{v}_n^2$.

In a vessel with fixed walls in thermal equilibrium, if the liquid background is flowing, its velocity $\mathbf{v}_s$ must have no component normal to the wall. Further, the total current normal to the wall must vanish, so that the normal component of $\mathbf{v}_n$ must also vanish at the walls. But in equilibrium $\mathbf{v}_n$ is constant everywhere and must therefore vanish everywhere. We say the normal fluid is stationary in equilibrium with fixed walls, even though the superfluid moves with the velocity $\mathbf{v}_s$. Incidentally, the superfluid velocity is irrotational, $\nabla\times\mathbf{v}_s=0$.

If the walls move together at constant velocity, then equilibrium results if $\mathbf{v}_n$ is this velocity; the normal fluid moves at the same velocity as the walls.

We extend Eq. (40) to situations slightly out of thermal equilibrium by assuming $\mathbf{v}_n$ is not constant but varies from place to place. The failure of equilibrium will bring in various irreversible processes associated with the normal fluid, such as viscosity. If we leave these out of account, the remainder of the hydrodynamic equations which result can be derived in exactly the manner already given by Dingle.[5] We need enter no further in this direction, as nothing new is gained. The resultant hydrodynamical equations can most easily be interpreted from the model of helium as consisting of two interpenetrating fluids.

Nevertheless, it is difficult to understand these partial fluids from a detailed kinematic point of view. Kinematically we have a general, or background fluid in which excitations move. The velocity of the superfluid, $\mathbf{v}_s$, is the general velocity of this background, but the density of superfluid is not ρ_0. The velocity of the normal fluid, $\mathbf{v}_n$, appears as a parameter in the distribution function. It can be shown to be the average group velocity of the excitations. But the difficulties arise if one tries to interpret the formula (42) for ρ_n from a direct kinematical point of view. It is not the average value of any quantity that can reasonably be ascribed to an individual excitation. It appears to have meaning only for the entire group of excitations in, or near, thermal equilibrium.

The division into a normal fluid and superfluid, although yielding a simple model for understanding

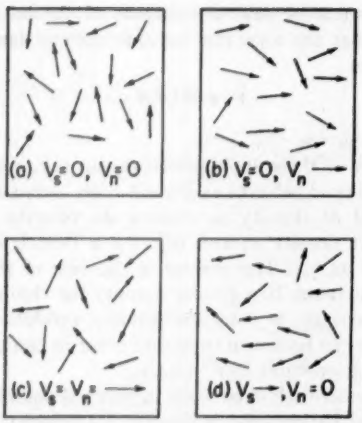

FIG. 4. At intermediate temperatures the main excitations are rotons which carry an intrinsic momentum, indicated by the arrows. If they drift relative to the background fluid, they tend to polarize upstream. This is illustrated for various values of v_s the background fluid velocity, and v_n the absolute drift velocity of the rotons. The total current is $\rho_0 v_s$ plus the polarization current of the rotons. Although mathematically correct, the separation of this current into the two parts, $\rho_s v_s$ and $\rho_n v_n$, characteristic of the two fluid model seems somewhat artificial from the microscopic viewpoint.

the final equations, appears artificial from a microscopic point of view. This opinion is shared by Landau and by Dingle.[5]

It is interesting to look at what is happening on a microscopic scale for various conditions of the velocities $\mathbf{v}_s$, $\mathbf{v}_n$. Consider a temperature not too low so that the predominant excitations are rotons. If the fluid is at rest a roton created with exactly the minimum energy Δ has no group velocity, but it has a momentum, of magnitude p_0 pointing in some direction. We will call it the direction of polarization and represent our roton by an arrow in this direction in Fig. 4. Not all rotons have exactly this energy Δ, but may differ by order kT from it, and have therefore nonzero group velocity. (Incidentally, the group velocity is parallel to or opposite to the polarization.) The rotons therefore may move about in a random manner like the molecules of a gas. Like gas molecules they can also have an average drift velocity relative to the fluid. Now we will assume [as required by (39), (41)] that if the rotons are drifting relative to the fluid in a certain direction, they tend (as a result of collisions among themselves) in equilibrium to polarize themselves in the direction in which they drift (that is, opposite to the velocity of the liquid moving past them). The general drift velocity of the rotons in space we call the normal fluid velocity, v_n. The motion of the fluid as a whole we call the superfluid velocity, v_s. Let us consider some examples.

First, with the fluid at zero velocity ($v_s=0$), and no drift of the rotons ($v_n=0$), they remain unpolarized (Fig. 4 (a)). If they are drifting to the right ($v_n>0$) they will tend to polarize in this direction, lining up to oppose the liquid passing them (Fig. 4 (b)). Now if the liquid is in motion ($v_s=v$), and all the rotons drift with the liquid in the same direction ($v_n=v$), there is no relative motion and no tendency to polarize (Fig. 4 (c)). This situation is not in equilibrium with stationary walls. Collisions of the rotons with the walls will stop their drifting motion and they will remain at rest relative to the walls. However, they will become polarized opposite to the fluid passing them (Fig. 4 (d)).

This interpretation of the velocities of the partial fluids of the two-fluid model is fairly simple and direct. It is otherwise with the current. The natural way to discuss the current (or momentum) from the microscopic view is to split it into two parts. First the current produced by the flow of the moving fluid, $\rho_0 v_s$, and second the current produced by the polarization. It is then easy to see what the current is in each case. In the first two cases, 4a, 4b, we have no general fluid motion so that the current is all due to polarization, zero in case *a*, to the right in case *b*. If the fluid moves, but the rotons remain unpolarized, as in 4c, the current is $\rho_0 v_s$, purely due to liquid motion. When the rotons polarize, as in 4d, to oppose this background motion the total current is reduced. But this natural separation is not the same as that utilized in the two fluid view. It is difficult to identify, for example, what is called the current of the normal fluid. It is not current carried by the rotons as they drift from one place to another (which they do with the velocity v_n) because the roton as such carries no mass but is only a disturbance in the liquid. Certainly the value of ρ_n would be hard to obtain this way because the rotons contribute to the current mainly by their polarization, and not by their drift motion. (Of course in a case such as 4b the polarization is in the direction of the drift so we could say the drift acts *as if* it carries current, because it induces polarization. ρ_n is then the ratio of the polarization current to the drift velocity which produces it.)

On the other hand, it is evident that the entropy flow is produced entirely by the drift motion (and not the polarization) of the rotons. Hence it is easy to see why all the entropy flows with the velocity v_n.

THE CONSERVATION OF CURRENT

In the last section we considered a packet of solutions (2) and found that we could picture an excitation in the form of a packet carrying a total current $\hbar\mathbf{k}/m$, and drifting at a group velocity $\partial E/\partial \mathbf{p}$, but as not appreciably altering the density. But such a picture is inconsistent with the conservation of matter. To take an extreme example, for a roton of the minimum energy Δ the group velocity $\partial E/\partial \mathbf{k}$ is zero, but the current $\hbar k_0/m$ is large. If such a current is distributed over a finite region in such a way that the direction is everywhere the same, we evidently cannot conserve material.

On the other hand, it is well known that one can

demonstrate the conservation of matter,

$$\partial\rho(\mathbf{a})/\partial t=\nabla\cdot\mathbf{j}(\mathbf{a}), \tag{44}$$

from the Schrödinger equation. The reason that our wave function does not satisfy (44) is that it is not an exact solution of the wave equation. This shows an inaccuracy in our approximate wave function (17).

One way that suggests itself to resolve it, in the case of rotons with $k=k_0$, is to propose a superposition of two waves with opposite momenta $\mathbf{k}$ and $-\mathbf{k}$, like $g(\mathbf{r})\cos(\mathbf{k}\cdot\mathbf{r})$. In this case the current density is zero, and everything is all right. Furthermore, if the same small momentum $\mathbf{l}$ is given to each, so the momenta become $\mathbf{k}+\mathbf{l}$ and $-\mathbf{k}+\mathbf{l}$ with $k=k_0$, the drift velocity $\partial E/\partial\mathbf{k}$ is the same for each partial wave, so that the packet stays together.

On the other hand, with stronger collisions with walls and phonons perhaps the two momentum components would become separated. Further $g(\mathbf{r})\sin(\mathbf{k}\cdot\mathbf{r})$ is just as good a solution, and it must have almost exactly the same energy even if interactions are taken into account, because the exact position of the nodes in a large packet cannot be important. Therefore, a linear combination must again be a possibility and we are led back to the exponential, and to the difficulty of current conservation. This lack of conservation is a symptom that all is not too well with our wave function. It is true that in the cosine case the symptom is hidden, but the conclusion should stand that the wave function could be improved.

The problem can be resolved by considering more complicated functions representing interaction of the excitation with the flow of fluid in its surroundings. One way the current could be conserved would be to have a general return flow of fluid in the region outside the packet. We therefore try the solution

$$\psi=\textstyle\sum_i g(\mathbf{r}_i)\exp(i\mathbf{k}\cdot\mathbf{r}_i)\exp[i\sum_j s(\mathbf{r}_j)]\phi, \tag{45}$$

with the hope of finding an s which produces a velocity distribution $\mathbf{v}=m^{-1}\nabla s$ which shows such a reverse flow. Let us first consider such a packet in otherwise stationary liquid. Then as a boundary condition s should go to zero as we go far from the packet. Substitution into the variation integral gives (32). For the current and density we use our approximations, that $\mathbf{j}(\mathbf{a})$ is given by (36), and $\rho(\mathbf{a})=\rho_0$. There results

$$\mathcal{E}/g=E(k)+\int\mathbf{j}(\mathbf{r})\cdot\nabla s(\mathbf{r})d^3\mathbf{r}+\frac{\rho_0}{2m}\int\nabla s\cdot\nabla s d^3\mathbf{r}, \tag{46}$$

where we have put, for the packet energy, $E(k)$, which is nearly correct. Variation of s to find a minimum gives the equation

$$\nabla\cdot(\mathbf{j}+\rho_0\nabla s)=0. \tag{47}$$

This equation determines s if we impose the boundary conditions $s\to 0$ far from the packet. Call this solution s_0 and the velocity distribution $\mathbf{v}_0=\nabla s_0/m$. It is like the field produced by the charge density $\nabla\cdot\mathbf{j}$, that is, at large distances the field of a dipole. It represents the back flow expected. Furthermore the total current operator has for our function the value

$$\mathbf{J}_0=\mathbf{j}+\rho_0\nabla s_0 \tag{48}$$

so that (47) says that now the total current is conserved.

There is a small shift in energy. Substitution of (47) into (46) gives the extra energy (reduction)

$$-\frac{\rho_0 m}{2}\int\mathbf{v}(\mathbf{r})\cdot\mathbf{v}(\mathbf{r})d^3\mathbf{r}.$$

If the order of the dimensions of the packet are L, the current $\hbar\mathbf{k}/m$ is distributed over a volume L^3, so the velocities are of order $\hbar k/mL^3$ and the kinetic energy $(\rho_0\hbar^2k^2/mL^6)L^3$ varies as $1/L^3$. But to confine the packet to such a dimension wave numbers of order $1/L$ in $g(r)$ must be used, so we find from (18) (for the case $k=k_0$) excess energies of order $1/L^2$ needed to confine the packet. Thus, for large packets, spreading the packet over even larger dimensions will decrease the energy, in spite of the energy of the currents we have just calculated. For extremely small packets our analysis does not hold because of the approximations made.

Here we have just gone far enough to save the theorem of conservation of current. We have only dealt with the background current in a semiclassical way. More complex states consisting of superpositions of expressions like (45) should be considered if a correct calculation of the quantum-mechanical "self-energy" of a roton due to coupling with the general velocity field is to be carried out. Since the "self-energy" is negative, the corrected value of Δ will be nearer the experimental result. This problem is being studied.

A more correct picture of a packet excitation, then, is that of a kind of region of polarization (that is, $\mathbf{j}$) which induces a distribution of velocity field around it, $\rho_0\nabla s_0$. The field is analogous to that produced by electrical polarization, the electric field $\mathbf{E}$ corresponding to the velocity field $\nabla s/m$, and the electric displacement vector $\mathbf{D}$ being analogous to the total current density, since its divergence vanishes. We can use this analogy to determine the behavior of the system directly, as it is easily verified that all the equations correspond. There is, however, one important difference. The signs of interaction are reversed. Thus (39) shows that rotons tend to line up opposed to the external field $\mathbf{v}$, while electric dipoles line up with the field. The analogy must therefore be completed with the remark that the rotons are dipoles of a gravitational type; that is, like poles attract, unlike repel.

The energy of a dipole in an external field is still the moment times the external field, even though the dipole itself creates some field of its own. None of the conclusions of the previous section are changed, therefore.

ROTON INTERACTION VIA THE VELOCITY FIELD

The fact that one roton creates a velocity field in which another may interact produces a kind of interaction between rotons. This is possibly one of the major sources of interaction, especially for not too high roton density. It is interesting to try to see what effect it has. Suppose we consider rotons as small packets all separated from one another and acting as dipoles of strength $\mathbf{p}/m$. Let us suppose there is an average polarization $\mathbf{P}$ per unit volume, and an average background velocity $\mathbf{v}_s$. The mean current density is then

$$\mathbf{J}=\rho_0\mathbf{v}_s+\mathbf{P}. \tag{49}$$

The actual field at any point is not $\mathbf{v}_s$ because of the local variations produced by the individual dipoles. Call the $\mathbf{w}$ velocity that an average dipole feels both from the average effect $\mathbf{v}_s$ and from its neighbors. The latter contribution is proportional to the polarization. In fact, as Lorentz showed for dipoles in random positions, it is $\frac{4}{3}\mathbf{P}/m$, hence

$$\mathbf{w}=\mathbf{v}_s+\alpha\mathbf{P}\rho_0^{-1}, \tag{50}$$

where[11] $\alpha=\frac{4}{3}$. The case $\alpha=0$ is the case previously studied which neglects direct effects between the dipoles.

The energy of a roton in this field is $E(\mathbf{p})+\mathbf{p}\cdot\mathbf{w}$. The statistical mechanics will then be governed by the function,

$$f=kT\int \ln\{1-\exp[-\beta(E(\mathbf{p})+\mathbf{p}\cdot\mathbf{w}-\mathbf{p}\cdot\mathbf{u})]\}d^3\mathbf{p}/(2\pi)^3. \tag{51}$$

Here $\mathbf{u}$ is zero for equilibrium with fixed walls, and is the normal fluid velocity, $\mathbf{u}=\mathbf{v}_n$. The average polarization then is given by

$$\mathbf{P}=m^{-1}\partial f/\partial\mathbf{w}. \tag{52}$$

The internal energy of the system is

$$U=\tfrac{1}{2}m\rho_0\mathbf{v}_s^2+m\mathbf{P}\cdot\mathbf{w}-(\alpha m/2\rho_0)\mathbf{P}^2+\langle E(\mathbf{p})\rangle, \tag{53}$$

where the average value of $E(\mathbf{p})$ is

$$\langle E(\mathbf{p})\rangle=\int E(p)\{\exp[\beta(E(p)+\mathbf{p}\cdot(\mathbf{w}-\mathbf{u}))]-1\}^{-1}d^3\mathbf{p}/(2\pi)^3$$
$$=f+TS-(\mathbf{w}-\mathbf{u})\cdot m\mathbf{P}.$$

The entropy is $S=-\partial f/\partial T$.

Expanding up to second order in the velocities one finds that the current can still be written as $\rho_s\mathbf{v}_s+\rho_n\mathbf{u}$, and the excess internal energy (at constant entropy and total current) as $\frac{1}{2}\rho_s\mathbf{v}_s^2+\frac{1}{2}\rho_n\mathbf{u}^2$, provided that one writes $\rho_s=\rho_0-\rho_n$ and

$$\rho_n=\rho_n^0(1+\alpha\rho_n^0/\rho_0)^{-1}, \tag{54}$$

where ρ_n^0 is the old expression (42) valid for the case $\alpha=0$. Therefore the expression for the velocity of second sound,

$$c_2=\left(\frac{\rho_n}{\rho_s}\frac{TS^2}{c_v}\right)^{\frac{1}{2}},$$

is unaltered when expressed in terms of ρ_n, etc. Only the theoretical formula for ρ_n is slightly modified. However, the modification is appreciable only when ρ_n^0/ρ_0 is not small, that is, near the transition. At the transition where c_2 goes to zero, ρ_n must equal ρ_0 so that ρ_n^0 given in (42) must equal $1/(1-\alpha)$ or 1.5. Actually ρ_n^0 varies very rapidly in this region so this makes no appreciable change in evaluating Δ and p_0 from the data. Furthermore in this region there may be other interactions which should alter our statistical mechanical analysis anyway. (Actually we cannot even be sure that rotons act as small individual dipoles until we have improved the wave function to include the interaction with the velocity field, as a quantum field.)

With interacting dipoles we would expect the analogue of a transition corresponding to the Curie point for electric dipoles. The analog of the condition for the Curie point comes out to be exactly the criterion that the expression (54) for ρ_n (with ρ_n^0 substituted from (42)) becomes equal to ρ_0. There are a few surprises here, though. Firstly, ordinarily the Curie transition occurs as we lower the temperature, but here it appears on raising the temperature. That is because the Curie point depends markedly on the density. Dipoles polarize if they are cold and dense. In our case at low temperatures they are cold, but not dense enough. As the temperature rises, the density does also, very rapidly, until a point is reached where spontaneous polarization appears even though the temperature has been raised. Another surprise is the fact that there is a transition even if the local field effect is neglected ($\alpha=0$). This difference is a result of the change of sign of the forces. Our dipoles polarize most easily if arranged in a flat region, while for electrical dipoles a needle-like region is preferred. With all the dipoles polarized parallel in the sheet the outside field is zero if the internal field opposed the polarization. But this opposition of polarization and field is just the stable condition for the rotons. In fact, the mutual local field α tends to *de*polarize them and *raises* the transition temperature. (For Onsager's value of α, (54) shows no transition for any temperature.)

One might be tempted to speculate that we could carry the statistical mechanical analysis right up to the transition point and beyond, by simply assuming that the only interaction of importance among rotons is the coupling with the general velocity field. One would just hope that other interactions or limitations to the number of degrees of freedom are not as important as

[11] Onsager has shown that if one deals with permanent dipoles, mutually impenetrable and roughly spherical, this value of α is in error. If his analysis applies to our case, the value $\alpha=\rho_0(\rho_0+2\rho_s)^{-1}$ results [L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936)].

one would otherwise guess. Aside from the amusing twist that helium I would then be the polarized, organized state, serious difficulties arise. One can analyze these things from the statistical formulas, if the velocities are not considered small and are not expanded. One can, without loss of generality, take states of total current zero, and for simplicity take $\alpha=0$. What happens is this. For any temperature below the transition there are two equilibrium states possible, one unpolarized with $v_s=0$, and the other polarized at finite v_s. Since the latter has higher free energy it is an unstable equilibrium, but the $v_s=0$ is stable (actually only metastable[12]). As we approach the transition point the polarization of the unstable state approaches zero. Above the transition point (more correctly, the point when $\rho_n=\rho_0$) only the $v_s=0$ state is in equilibrium and that is unstable. The instability arises this way. There is a high density of rotons. If a little polarization develops, their energy is reduced. This increases the number of rotons in equilibrium at a fixed temperature as well as the polarization, so that if there is no limit to the number of rotons there is no stable state.

On the other hand, if a limitation of roton number such as (28) is imposed, stable polarized states exist at the higher temperature. But the transition to that state occurs as a first-order transition, and there is another transition at still higher temperatures when the polarization disappears again.

It is therefore evident that we do not correctly describe the region very close to the transition by the usual energy expression (53) (with or without $\alpha=0$). The interactions between rotons is playing a more complicated role than (53) can describe.

In a previous paper an expression was given for the partition function which was presumably reasonably satisfactory right across the transition. However, the analysis was too difficult to carry out. Now that a more detailed picture of the behavior below the transition is available it may be easier to see how that expression can be treated. We still lack a clear picture of what happens in the few tenths of a degree on either side of the λ point.

INTERACTIONS BETWEEN EXCITATIONS

Interactions between the excitations will lead to various irreversible processes, such as viscosity, attenuation of second sound, etc. These questions have been studied by Landau and Khalatnikov.[13] The most important factor in the various mean free paths which are involved is the change in density of the phonons and rotons with temperature. The absolute cross sections depend on the details of the interactions. Interactions between phonons can be thought of as arising from a nonlinear equation of state. An interaction between a phonon and a roton would result if rotons have a different energy for different pressures. According to the theory presented here their energy is $\hbar^2k^2/2mS(k)$. If the liquid is compressed k^2 increases. On the other hand $S(k)$ probably increases even more rapidly from the increase in local order produced by squeezing the nearly impenetrable atoms into a smaller space. Therefore we expect Δ to decrease with pressure. This provides a mechanism for roton-phonon interaction. It also has other effects. The presence of a roton would cause a small increase in density in its neighborhood with the effect falling off inversely as the distance from the roton. This provides a mechanism of long-range interaction between rotons in addition to that due to coupling with the general velocity field. The roton-roton interaction at short distances is a more difficult problem, which probably cannot be adequately solved until a more accurate wave function is available for the roton state.

If the roton energy Δ decreases when the liquid density increases then we would expect that in equilibrium the liquid would shrink if the number of rotons is increased. The volume decrease as the λ point is approached is probably a consequence of this effect. The fall of the λ temperature with rising pressure is thermodynamically related. It may also be seen directly from (42) (supposing the λ point to be $\rho_n=\rho_0$) considering that Δ decreases as the density rises.

SUPERCONDUCTIVITY

It has been suggested that superconductivity is analogous to superfluidity. What can we learn of the former from our study of the latter? It is interesting that if the He atoms were charged (and their net charge canceled by a uniform fixed background charge of opposite sign) the liquid would imitate many of the features of a superconductor. In a magnetic field at absolute zero the London[14] equation, $\mathbf{j}=e\mathbf{A}/mc$, would hold. This is because stirring of atoms is equivalent to interchange so that in the lowest state the wave function cannot vary if atoms are stirred, and the part of the current depending on the wave function gradient vanishes. Other states would take a finite energy to create, there would be states of permanent circulation in multiply connected rings, there would be a second-order transition, etc. How this close analogy is to be interpreted is not clear. The things which, it has been argued here, apply physically to helium cannot be justifiably taken over to the case of Fermi particles, or such particles in interaction with lattice waves, without a complete investigation of their validity in the new environment. For example, at present there is not, in the author's opinion, justification for assuming

[12] That is, the conventional free-energy expressions for arbitrary v_s are, strictly speaking, not self-consistent. For any temperature there is always some value of v_s for which the free energy is less than its value for $v_s=0$.

[13] L. Landau and I. Khalatnikov, J. Exptl. Theoret. Phys. (U.S.S.R.) **19**, 637, 709 (1949).

[14] F. London, *Superfluids* (John Wiley and Sons, Inc., New York, 1950), Vol. I.

that the form (2) is a reasonable wave function for an assembly of Fermi particles, with ϕ the ground-state function for such particles. In fact, there are definite arguments against it. Possibly the close analogy should only be used to tell us what the problem of superconductivity is. It is, from this point of view, the problem of showing that, in the metal, aside from phonons there are no (or only very few) states of very low energy just above the ground state.

DISCUSSION

We still have left unsolved at least three basic questions. One is to find a clear description of the neighborhood of the transition. A second is to obtain a more perfect roton wave function. The third is to describe states for which the superfluid velocity is not vortex-free. So far we have $\nabla\times\mathbf{v}_s=0$. At high velocities when more energy is available, more complicated motions might be excited. The evidence of high resistance to capillary flow at the higher velocities indicates this. A new element must presumably be added to our picture. We hope to publish some views on this third problem at a later time.

We have limited ourselves to a qualitative analysis of the more curious features of the behavior of liquid helium. The problem of obtaining $S(k)$ or the correlation function for the ground state quantitatively from first principles is beyond the scope of this work.

It has been argued[1] that He^3 atoms in low concentration, in He^4 would act as a gas of free particles, but with an effective mass m'' higher than that of one atom. This mass m'' is calculated in the appendix, where it is found to be about six atomic mass units.

The author has profited from conversations with R. F. Christy and with Michael Cohen.

APPENDIX

According to II an impurity atom of He^3 (at infinite dilution) should behave as an essentially free particle except that its effective mass m'' should exceed the true mass of He^3 due to the inertia of the He^4 atoms which must make way for it as it moves. We shall calculate this excess mass here. First we suppose the impurity atom had the same mass m as the other He^4 atoms (i.e., we neglect the difference in mass of He^3 and He^4).

The wave function for such an atom (coordinates $\mathbf{r}_A$) moving with momentum $\hbar\mathbf{k}$ might be conjectured to be $\exp(i\mathbf{k}\cdot\mathbf{r}_A)\phi$, where ϕ is the ground state of the system (which is the same as if all the atoms were identical). This, as a trial function, gives the variational energy as $\hbar^2k^2/2m$ and shows no mass correction. It does not represent with sufficient accuracy the other atoms moving back when atom A moves forward. This suggests the trial function

$$\psi=\exp(i\mathbf{k}\cdot\mathbf{r}_A)\exp[i\textstyle\sum_i s(\mathbf{r}_i-\mathbf{r}_A)]\phi, \qquad (1\text{-a})$$

where the velocity field $\nabla s(\mathbf{r}_i-\mathbf{r}_A)$ represents a backflow which depends only on the distance from the impurity. To omit the term $i=A$ in the sum we take $s(0)=0$. Substitution into the variational principle gives

$$\mathcal{E}=(\hbar^2/2m)\int[k^2-2\mathbf{k}\cdot\textstyle\sum_i\nabla s(\mathbf{r}_i-\mathbf{r}_A)+\sum_i\sum_j\nabla s(\mathbf{r}_i-\mathbf{r}_A)\cdot\nabla s(\mathbf{r}_j-\mathbf{r}_A)+\sum_i\nabla s(\mathbf{r}_i-\mathbf{r}_A)\cdot\nabla s(\mathbf{r}_i-\mathbf{r}_A)]\rho_N d^N\mathbf{r}, \qquad (2\text{-a})$$

with $\mathcal{I}=1$, as ψ is normalized. We write $\rho_2(\mathbf{r}_i,\mathbf{r}_A)=\rho_0 p(\mathbf{r}_i-\mathbf{r}_A)$, and $\rho_3(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_A)=\rho_0 p_3(\mathbf{r}_i-\mathbf{r}_A,\mathbf{r}_j-\mathbf{r}_A)$, and measure all distances from the point $\mathbf{r}_A$, so that (2-a) becomes

$$2mh^{-2}\mathcal{E}=k^2-2\mathbf{k}\cdot\int\nabla s(\mathbf{r})p(\mathbf{r})d^3\mathbf{r}+\int\int\nabla s(\mathbf{r})\cdot\nabla s(\mathbf{r}')p_3(\mathbf{r},\mathbf{r}')d^3\mathbf{r}d^3\mathbf{r}'+\int\nabla s(\mathbf{r})\cdot\nabla s(\mathbf{r})p(\mathbf{r})d^3\mathbf{r}. \qquad (3\text{-a})$$

Then s is to be chosen to minimize this expression.

The function $p_3(\mathbf{r},\mathbf{r}')$ is the probability of finding one atom at $\mathbf{r}$ and another at $\mathbf{r}'$ if there is an atom at the origin ($p(\mathbf{r})$ is just the probability of finding one at $\mathbf{r}$ if one is at the origin). We do not know what this function p_3 is, but in this problem we can approximate it by $p(\mathbf{r})p(\mathbf{r}')$ (except at the origin). A given atom is surrounded by many (eight or ten?) nearest neighbors, and the distribution along one radius r and another r' must be nearly independent except for the relatively small solid angle where the atoms at $\mathbf{r}$ and $\mathbf{r}'$ are close together. Even here, if the functions $s(\mathbf{r})$, $s(\mathbf{r}')$ are smooth enough the average of p_3 over such angles may still give nearly the same result as $p(r)p(r')$. With this substitution our problem is that of minimizing

$$2mh^{-2}\mathcal{E}=\left[k-\int\nabla s(\mathbf{r})p(\mathbf{r})d^3\mathbf{r}\right]^2+\int\nabla s(\mathbf{r})\cdot\nabla s(\mathbf{r})p(\mathbf{r})d^3\mathbf{r}. \qquad (4\text{-a})$$

The error made by this approximation is

$$\Delta\mathcal{E}=(\hbar^2/2m)\int\int\nabla s(\mathbf{r})\cdot\nabla s(\mathbf{r}')[p_3(\mathbf{r},\mathbf{r}')-p(\mathbf{r})p(\mathbf{r}')]d^3\mathbf{r}d^3\mathbf{r}'. \qquad (5\text{-a})$$

The $s(\mathbf{r})$ which minimizes (4-a) behaves as a dipole field (as $\mathbf{k}\cdot\mathbf{r}/r^3$) at large distances and this produces some convergence difficulties in the first integral in (4-a). They may be easily straightened out as follows. The minimum energy is only very slightly altered if the function $s(\mathbf{r})$ is altered at very large distances in such

a way that it falls off eventually more rapidly than $1/r^2$. For such a function the integral may be done by parts, the integrated part vanishing, so we may write

$$2mh^{-2}\mathcal{E}=\left(k+\int s(\mathbf{r})\nabla p(\mathbf{r})d^3\mathbf{r}\right)^2 + \int \nabla s(\mathbf{r})\cdot\nabla s(\mathbf{r})p(\mathbf{r})d^3\mathbf{r}. \quad (6\text{-a})$$

But in this form all the integrals have a definite limit even if $s(\mathbf{r})$ has no convergence factor (and therefore varies as $1/r^2$). We may therefore use (6-a) and avoid ambiguities from conditionally convergent integrals.

The variational principle shows that s must be a solution of

$$\nabla\cdot[p(\mathbf{r})((1-\beta)\mathbf{k}-\nabla s(\mathbf{r}))]=0, \quad (7\text{-a})$$

where we have set

$$-\int s(\mathbf{r})\nabla p(\mathbf{r})d^3\mathbf{r}=\beta\mathbf{k}. \quad (8\text{-a})$$

Multiplication of (7-a) by $s(\mathbf{r})$ and integration, using (8-a), tells us further that

$$\int \nabla s(\mathbf{r})\cdot\nabla s(\mathbf{r})p(\mathbf{r})d^3\mathbf{r}=\beta(1-\beta)k^2,$$

so (6-a) says

$$\mathcal{E}=\frac{h^2k^2}{2m}[(1-\beta)^2+\beta(1-\beta)]=\frac{h^2k^2}{2m}(1-\beta),$$

and the effective mass is $m/(1-\beta)$, an increase over m of

$$\Delta m=\beta m/(1-\beta). \quad (9\text{-a})$$

If the direction of $\mathbf{k}$ is taken as the z axis, the solution of (7-a) may be written in the form,

$$s(\mathbf{r})=(1-\beta)k(z-zv(r)/r), \quad (10\text{-a})$$

where $v(r)$ is a function of radius $r=(\mathbf{r}\cdot\mathbf{r})^{\frac{1}{2}}$ only, satisfying

$$\frac{d}{dr}\left(r^2p(r)\frac{dv}{dr}\right)=2pv, \quad (11\text{-a})$$

and such that v approaches r at large distances. This is easily solved numerically. We used the values of $p(r)$ determined by Reekie and Hutchison.[4] Starting for small r where $p(r)=0$, so that $dv/dr=0$, we chose v at some convenient value and integrated out to radii so large that $p(r)$ was effectively its asymptotic value ρ_0. Asymptotically v has the form $c(r+B/r^2)$ where B, c are constants. Since the equation is homogeneous we may divide the entire solution by c to obtain one with the correct asymptotic form. By substitution of (10-a) into (8-a) one can show, using (11-a), that the expression for (9-a) can be written

$$\Delta m/m=4\pi\rho_0B-1, \quad (12\text{-a})$$

where we have used the fact that

$$\int_{0+}^{\infty}(\rho_0-p(r))4\pi r^2dr=1 \quad (13\text{-a})$$

[the origin, where $p(r)$ has a δ function being excluded in this integral]. Actually it is difficult to obtain accuracy with this method because the asymptotic form of v is sensitive to the values of $p(r)$ used. Those of Reekie and Hutchinson[4] extended only up to $r=6$A and had to be taken from a graph, so that (13-a) was not accurately satisfied without some small arbitrary readjustments of the values.

On the other hand, the solution showed that $s(\mathbf{r})$ was nearly proportional to z/r^3. Since (6-a) is a variational principle we can therefore obtain a good value of the energy much more simply. We substitute the trial function,

$$s(\mathbf{r})=Az/r^3, \quad (14\text{-a})$$

directly into (6-a) and determine the parameter A to minimize $\mathcal{E}$. This gives

$$\Delta m/m=\tfrac{1}{2}\left(\frac{4\pi}{3}\rho_0\right)^2\left[\int_{0+}^{\infty}r^{-6}p(r)4\pi r^2dr\right]^{-1}, \quad (15\text{-a})$$

or, with the data of reference 4, $\Delta m=0.70m$, or 2.8 atomic mass units. (The numerical solution of the differential equation gave the same result within its accuracy of about 10 percent.)

It is difficult to evaluate the small correction arising from the term $\Delta\mathcal{E}$ of (5-a), for p_3 is unknown. If the atoms locally are nearly on a lattice, say face-centered, or body-centered, of cubic symmetry, $\Delta\mathcal{E}$ vanishes with the trial function (14-a).

If the mass of the impurity atom is not four atomic units it is readily shown that Δm is unchanged, provided that the distribution $p(r)$ of atoms around the impurity is assumed to be unchanged. This is also expected physically for the extra mass is due to the motion of the He^4 atoms in the environment of the impurity. Therefore for a He^3 atom the effective mass should be 5.8 atomic mass units. The higher zero-point motion of a lighter atom changes $p(r)$ by pushing the neighbors farther away, thereby raising Δm a little (but this effect must be a fraction of a mass unit because the effective mass is almost certainly less than the 6.8 which it would be if the He^3 had the larger mass of 4 units).

This result is not in good agreement with the determinations which have been made from experiment, as summarized by Daunt.[15] These give values nearer 8 or 9 atomic mass units.

[15] J. G. Daunt, Advance in Physics (Phil. Mag. Supplement) 6, 209 (1952), particularly p. 258.

Photoconductivity in Gold-Germanium Alloys

ROGER NEWMAN
General Electric Research Laboratory, Schenectady, New York
(Received January 8, 1954)

Photoconductivity has been observed at 77°K and 22°K in *n*- and *p*-type gold-doped germanium. The low-energy thresholds for impurity photoconduction are compatible with the activation energies found by conventional electrical measurements. Quenching effects were found in some *n*-type samples. Prominent quench bands were found at 0.8 and 0.66 ev. The characteristics of the quenching were observed as a function of light intensity, applied voltage and temperature. An interpretation of the quenching data in terms of a hole trap model is suggested.

INTRODUCTION

THE investigations of Dunlap[1] have shown that gold acts as an electron acceptor in germanium. On the basis of his resistivity and Hall data, he suggests a model in which each gold atom is capable of accepting one or two electrons. The first excess electron is bound in a state which is 0.15 ev above the valence band, while the second electron is bound with an energy which is 0.2 ev below the conduction band. These energies are such that if the number of gold atoms is greater than half the excess of normal donors over normal acceptors, so that all the normal donor levels are emptied, then the samples will be heavily deionized even at liquid N_2 temperature. Dunlap's model suggested that a study of the infrared photoconductive response of gold-doped germanium might be of interest in that confirmatory evidence for the proposed energy level scheme might be forthcoming. This proved to be the case.[2] In addition, a "quenching" of the photoconductivity was observed. The latter was of interest in that it offered an opportunity for greater insight into quenching phenomena in general in a material for which the energy level scheme and impurity content is under much better control than for conventional photoconductive materials such as CdS.[3]

EXPERIMENTAL

Samples of gold-doped germanium were obtained from W. C. Dunlap, Jr., in the form of small bars averaging about 3×3×10 mm. These were cut from the original ingots in such a way as to attempt to maximize sample homogeneity. In some cases, the samples were those used for Hall and resistivity measurements. In most cases, the sample resistance exceeded 10 megohms at 77°K, although some measurements were made with low resistance samples as well. For *n*-type samples contacts were soldered variously of tin, or alloys of either gold-antimony or indium-arsenic. Tin or indium were used for *p*-type samples. Samples were glued to a special copper block with a calorimeter adhesive. Electrical insulation between the block and the sample was assured by a layer of "Formex" enamel baked on to the block. The assembly was inserted into a cavity of a cryostat fitted with NaCl windows.

Illumination from the exit port of the spectrometer was substantially uniform over the length of the samples. The contact regions were shielded from direct radiation in some runs and not in others. No important differences were found. Temperatures were measured by a Au, Co-Ag, Au thermocouple soldered to one contact point of the sample. The measured temperatures probably represent the highest temperatures of the sample.

The photoconductive measurements were made in the usual way with a series circuit consisting of the sample, a load resistor, and a battery. The photocurrent was detected as a change of voltage across the load. Measurements were made as follows: (a) In dc light, an Applied Physics Corporation, "Vibrating Reed Electrometer" was used as a detecting voltmeter. (b) In chopped light (13 cycles), a Perkin-Elmer thermocouple amplifier with a high input impedance preamplifier was used in conjunction with a chart recorder. (c) For studies of transient effects, an oscilloscope with a direct coupled input was used.

A Perkin-Elmer monochromator was used with NaCl and CaF_2 prisms. The energy distribution of the globar radiation source was checked against a Perkin-Elmer thermocouple. Conventional methods were used to monitor the stray light response of the samples at longer wavelengths.

RESULTS

Normal Photoconduction

The normal photocurrents are linear in applied voltage and incident intensity at least over the range studied. Figures 1 and 2 show the photoconductive response of *n*- and *p*-type gold-doped germanium samples. The curves shown are typical of those found for four samples of each type derived from different ingots. The curves were obtained from chopped light

[1] W. C. Dunlap, Jr., Phys. Rev. **91**, 1282 (1953).
[2] Results agreeing in part with those obtained here have been described by W. Kaiser and H. Y. Fan, Phys. Rev. **93**, 911(A) (1954).
[3] E. A. Taft and M. H. Hebb, J. Opt. Soc. Am. **42**, 249 (1952). This paper gives a detailed bibliography on quenching.

data. Curves obtained by dc techniques were essentially identical with those shown.

The structure in the vicinity of 0.35 to 0.4 ev is spurious and represents the combined effect of the absorption of the adhesive material on the back of the sample and to a lesser extent condensed contaminants (e.g., water vapor) on its front. The solid curves represent what is felt the true response should be.

The different curves shown in the figures were normalized by direct measurement of the dc photocurrents and the incident radiation power from the source at a photon energy of 0.5 ev. The run-to-run reproducibility of the reference value for a given sample was of order ±20 percent. Differences for a

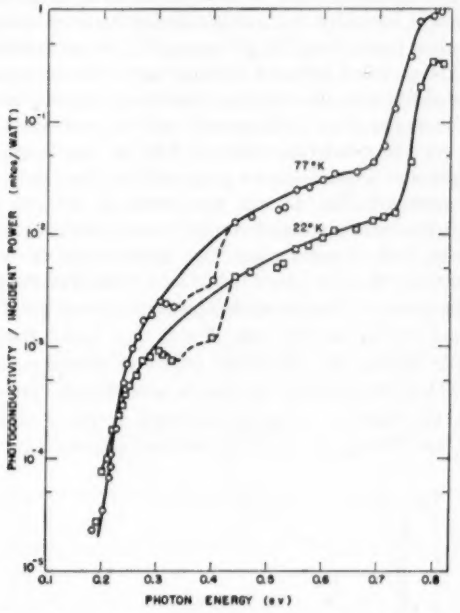

FIG. 1. Spectral response of the photoconductivity of *n*-type gold-doped germanium at 77°K and 22°K. Sample thickness 3 mm. Estimated $n_{Au}=4\times10^{14}$/cc.

given sample presumably resulted from irreproducible geometry.

Evidence for a complexity in the photoconduction is found from a study of time constants. Figure 3a shows the pattern of photoconductive response for an *n*-type sample at ~22°K to a light on–light off cycle. Radiation well in the impurity range (0.35 ev) was used. Clearly two processes with widely differing time constants are involved here. The rapid response is best described by a constant of order 10^{-3} sec, the slow response by a constant of about 0.5 sec. At 77°K the same time constants are found for this sample, but at this temperature the major part (~90 percent) of the photoresponse is described by the rapid response. When quenching effects are also present, the response pattern gets very complex. Figure 3b shows such a

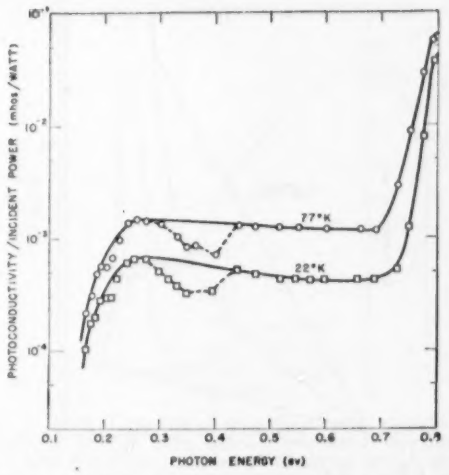

FIG. 2. Spectral response of the photoconductivity of *p*-type gold-doped germanium at 77°K and 22°K. Sample thickness 2.5 mm. Estimated $n_{Au}=8\times10^{14}$/cc.

pattern for the same sample when illuminated with 0.8-ev radiation. The quenching effects will be discussed at greater length below.

The complex response patterns were not found in all *n*-type samples. In some, no evidence was found for long response times except for quenching effects when these were present. Similarly, no long response times were found in any of the *p*-type samples studied. For the latter, decay times for the response to a pulse of light were of order 10^{-4} sec or less at 77°K, and about 10^{-3} sec at 22°K.

Quenching Effects

Early in the investigation it was found that quenching effects were present in the gold-doped samples. In

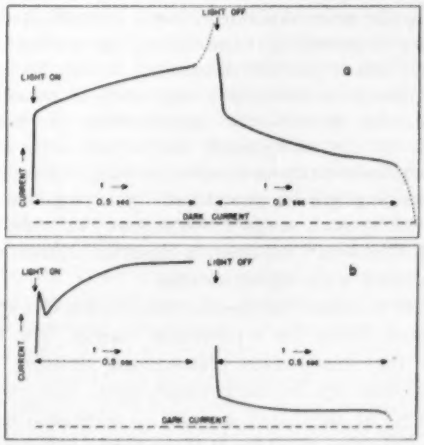

FIG. 3. A time-response pattern of an *n*-type sample at 22°K: (a) to 0.35-ev radiation; (b) to 0.80-ev radiation (from an oscillogram). The dotted sections result from nonlinearity in the sweep.

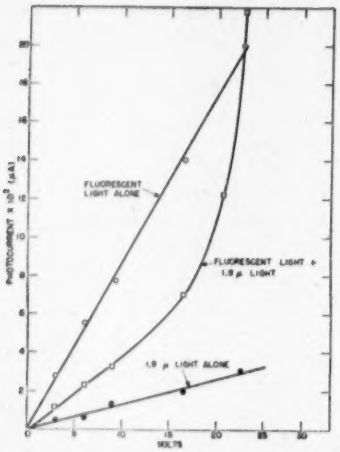

FIG. 4. Curve of dc photocurrent *vs* voltage for an *n*-type sample showing quenching; the effect of different modes of illumination are shown.

the usual definition, quenching is the effect by which a photocurrent and/or "dark" current through the specimen is reduced by illuminating the latter with another beam of light. Figure 4 illustrates the effect. It shows the steady-state photocurrent-voltage characteristic for a sample at 77°K when illuminated with a fluorescent lamp, when illuminated with 1.9μ radiation and when illuminated with the fluorescent light and the 1.9μ radiation. Parenthetically, it is worth noting that with the sample at 22°K under the same conditions of illumination, the quenching increased. The illustrated photocurrent could be decreased to about 10–20 percent of its initial value.

Figure 5a shows a typical cycle of response for a sample illuminated with quenching light. It is similar to other reported quench effects.[2] However, it has been observed that whether or not a sample shows a quench effect in the sense described above, depends upon the intensity of quench light and the voltage applied to the sample. This is partially illustrated in Fig. 5b, which shows that at a sufficiently high level of quench intensity, the steady-state photocurrent is positive relative to the background. To include such effects and to include examples of self quenching, the definition of quenching will be generalized by saying that it is exhibited by such samples which show, at some point in the "light-on" response, a negative derivative of conductivity with respect to time.

Figure 6a shows the spectral distribution of a system of quench bands for a particular sample. The figure shows the steady-state photocurrent relative to a photocurrent set by background light. The presence of a band at about 0.66 ev with an additional weaker effect extending to about 0.35 ev seems indicated. As there might be some question as to the shape of the high-energy side of the 0.66-ev band, an entirely different technique was used to observe the quench. Figure 6b shows the spectrum of the ratio, in percent, of the peak transient photocurrent (see Fig. 5b for definitions) minus steady-state photocurrent to the steady state photocurrent, a quantity termed percent "Quench," which should give a measure of the quenching effect. The measurements were made by an oscilloscopic technique. The monochromatic light level was adjusted so as to give the same steady-state photocurrent at each wavelength. It seems quite clear that a high-frequency limit for this quench does exist and that the major effect is properly defined in terms of a band centering at about 0.66 ev.

The spectrum illustrated in Fig. 6a, and in particular the 0.66-ev band, seem to be characteristic of *n*-type specimens for which the gold concentration is estimated to be less than about 1×10^{14} atoms/cc. In some cases, the 0.66-ev band is found without any other structure being observable. In one case, however, the prominent band was found at 0.54 ev with only a weak band at 0.66 ev. The quenching effect at 0.66 ev was found in samples cut from 5 ingots prepared to have the low gold concentration. It was not found in samples for which the estimated gold content exceeded this value. It was first thought that the appearance or non-appearance of these bands might be a sensitive function of the ratio of donor atoms to gold atoms but this seemed not to be the case. For it was found that a sample having the stipulated gold concentration, but for which the number of donors was slightly greater than the number of gold produced acceptor states, (i.e., low resistance at 77°K) showed a quench in the

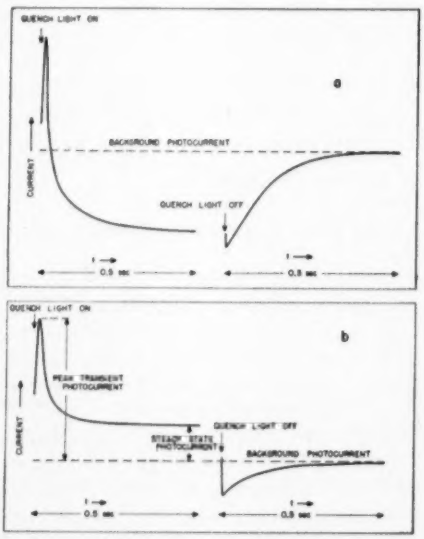

FIG. 5. A time-response pattern of a sample showing quenching when illuminated with quench light (1.9μ); the effect of different levels of quench intensity is shown in (a) and (b). The level in (b) is 3 times that in (a) (from an oscillogram).

same way as the high-resistance samples conventionally studied for which the ratio of the two concentrations is presumably reversed. It would, therefore, appear that the presence or absence of the 0.66-ev quench band is determined largely by the gold concentration.

In some n-type samples with gold concentrations higher than about 1×10^{14}, an example of self-quenching was observed. When light of appropriate wavelength in the fundamental region was turned on, the initial photocurrent decayed to a lower steady-state value. The curves observed with an oscilloscope resembled those of Fig. 5b. Since in this case, the radiation apparently produces an appropriate type of photocurrent as well as exciting the quenching, no additional background radiation source was required. Figure 7 shows spectral distributions of the percent "Quench" (defined above) for this quenching effect. Measurements, as in those displayed in Fig. 6b, were taken with the light level adjusted to give a fixed steady state photocurrent for each curve at all wavelengths. There appears to be a band centered at about 0.8 ev followed by a sort of continuum at energies higher than about 0.9 ev. The figure shows that the effect is decreased at increasing photoconductive levels. At a level about 5 times higher than the highest shown, no quenching was observable.

In other n-type samples with high gold concentration,

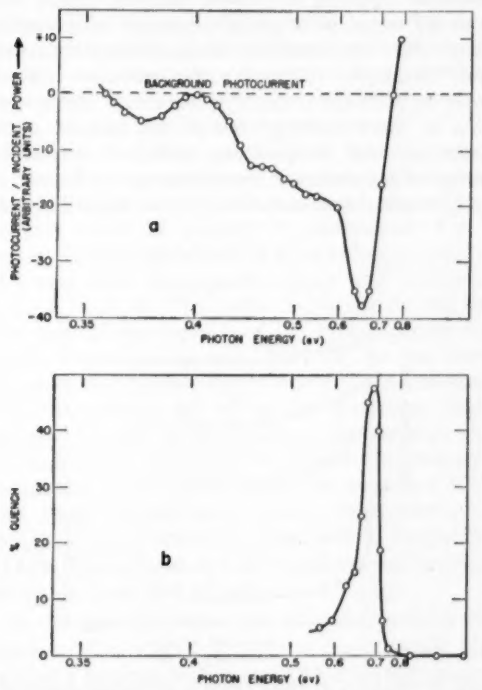

FIG. 6. Spectral response of an n-type sample showing a strong quenching at 0.66 ev: (a) measurement of dc steady state photocurrent relative to the background; (b) measurement using transients as described in text.

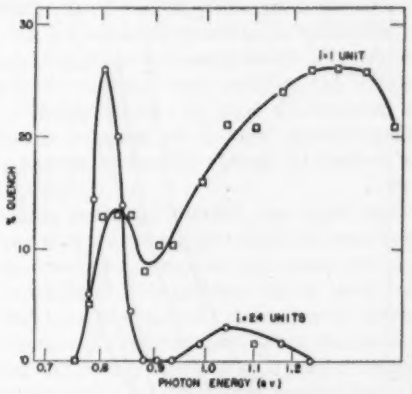

FIG. 7. Spectral response of an n-type sample showing self-quenching at 0.80 ev, using transient response as described in text. Effect of different levels of photocurrent (i) are shown.

the band at 0.8 ev was present but the continuum was too weak to be effectively measured. However, in all cases for which the 0.8-ev band was present, the 0.66-ev band was absent and vice versa. It was of extreme interest to note the apparent existence of a band at an energy exceeding the forbidden gap in germanium. The situation here is complex in that for this energy the radiation is not uniformly absorbed. Consequently, the possibility exists that a purely surface effect is being observed. The effect seemed, however, independent of surface treatment in that for a given sample the effect was present with a ground surface and with an etched surface.

It was also worth noting that at 22°K several samples showed slight quenching at energies exceeding about 0.85 ev which showed no corresponding effect at 77°K.

As was mentioned above, whether a sample shows a negative steady-state photoconductivity (relative to the background) depends upon the voltage applied to the sample and the amount of quenching radiation impinging on it. This is illustrated by Fig. 8 which shows a series of curves relating total dc photocurrent produced by a fluorescent light plus the quench light (1.9μ) to the intensity of the latter and to the voltage.

The complexity of these curves can be reduced to a certain extent. The response to quenching light is the sum of a normal photoconductive component linear with respect to voltage and intensity and a quenching component. The voltage and intensity behavior of the normal component may be determined by observing the response of the sample in the absence of any background light. An example of the current-voltage characteristic for the normal component is shown in the curve marked "1.9μ light alone" in Fig. 4. This normal component can then be subtracted from the total effect observed with the quench light plus the background photocurrent employed. One can then, presumably, observe the effect of the quench component alone,

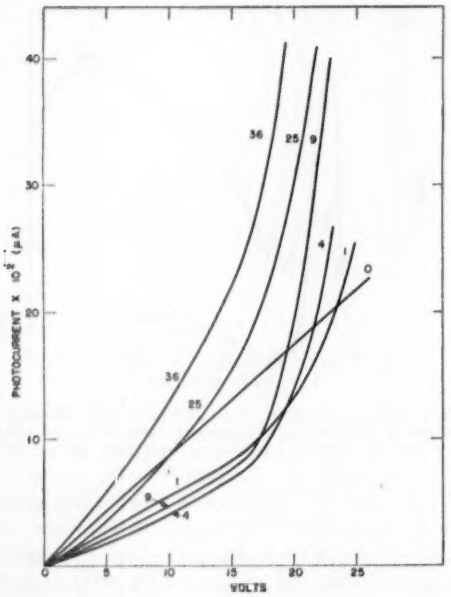

FIG. 8. Curves of dc photocurrent *vs* voltage for a constant level of background and different relative intensities of quench light (1.9μ) as indicated by the labels.

assuming that the two components are independent of one another. Data corrected in this way are shown in Figs. 9 and 10. Figure 9 shows the effect of the intensity of the quench component on different levels of background photocurrent. The ordinate, as in that of Fig. 10, is the total dc photocurrent minus the normal photocurrent produced by 1.9μ light at the same voltage and monochromatic light intensity. The apparent saturation of the quench effect with quenching intensity and the decreased fractional effect at higher background current levels are to be noted. Figure 10 shows the effect of different applied voltages on the quenching.

Ideally, the data of Figs. 9 and 10 should have been obtained with two monochromatized light beams. As this could not be readily arranged, a fluorescent lamp which has a negligible output in the quenching region was used as the exciting source. To insure that the effects were not surface phenomena, it was ascertained that essentially equivalent data were obtained when radiation reached the sample either directly or after passing through a germanium filter. Since only a limited range of background current could be obtained using a filter, the data of Fig. 9 were obtained without a filter.

The quenching process is characterized by a relatively slow decay of the photoconductivity from some peak value, when the quench light is turned on, down to a lower steady-state value. Similarly, when the quench light is turned off there is usually a negative dip followed by a slow rise to the conductivity set by the background. These transients seem describable as simple exponentials. The time constants at 22°K for a given sample for the two processes are illustrated in Fig. 11 as a function of the intensity of the quenching light (1.9μ). It is seen that the decay time in the quench light is a function of intensity whereas the recovery time in the "dark" (no quench light) is not. The "light on" decay time is also sensitive to the background level whereas the "dark" recovery time is not. For example, the former increased about a factor of two for a fourfold increase in background. It is also important to note that the dark recovery time, which seems a suitable descriptive parameter, only decreased by about a factor of two, in warming a sample from 22°K to 77°K. In some other samples "dark" recovery times of order 1 sec at 22°K were observed. The "dark" recovery time for samples showing the 0.8-ev quench were of order 0.1 to 0.5 sec at this temperature.

DISCUSSION

A satisfactory explanation of all the effects reported above, particularly those related to quenching, has not been forthcoming. However, it is felt that for heuristic purposes tentative hypotheses may be of value.

An analysis of the spectral dependence of normal photoconduction seems straightforward. The photoconductive response at photon energies, greater than about 0.7 ev, is, of course, the intrinsic photoconductivity. For the condition that (absorption coefficient)$^{-1}$>sample thickness, the response will be proportional to the absorption coefficient. The spectral shifts in the low energy tail of the intrinsic photoconduction with temperature, therefore, mirrors the corresponding change in the absorption coefficient.

At energies below about 0.7 ev, the impurity photo-

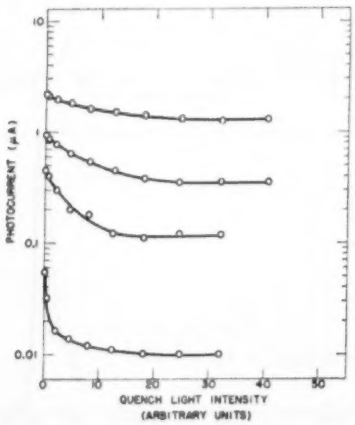

FIG. 9. Curves of dc photocurrent at a constant voltage (6 v) *versus* the intensity of quench light (1.9μ). The effects on different levels of background are shown. The effect of the normal component of response to the quench light has been subtracted (see text).

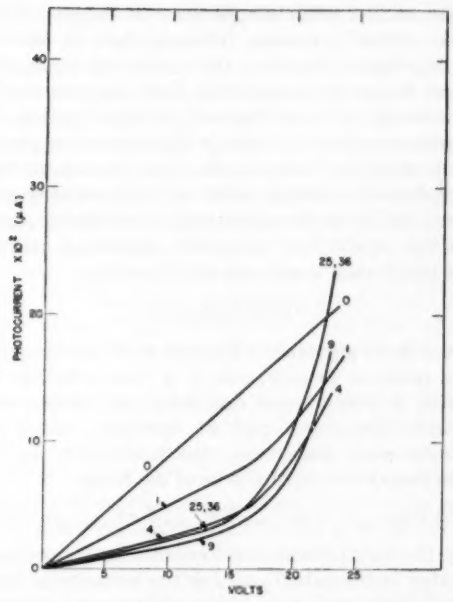

FIG. 10. Curves of dc photocurrent *vs* voltage at a given background level showing the effect of different relative intensities of quench light (1.9μ), as indicated by labels. The effect of the normal component of response to the quench light has been subtracted (see text).

conductivity is, presumably, dominant. Since the absorption coefficient, due to ionization of the impurity centers, may be estimated to be at most $10^{-1}-10^{-2}$ cm^{-1} (atom fraction$\times 10^6$), the photoresponse should give a measure of its spectral dependence.

Our results indicate a different spectral response for *n*- and *p*-type gold doped samples. Using Dunlap's model, it might be expected that the response of *p*-type samples would be simplest to understand. For the process in this case should involve only the ionization of a hole from the impurity center with a threshold at about 0.15 ev. This seems borne out by the data. The form of the curve is of the type expected for a simple ionization process. That is, as the energy decreases, the response should slowly rise to a maximum and then sharply fall off as the ionization limit is reached. There is an uncertainty in the point at which the ionization limit should be defined. By analogy to the intrinsic case, a point where the response is 10^{-3} of peak might be arbitrarily chosen. Extrapolating our data, it is found that this corresponds to an energy of 0.13 ev. The response at 0.15 ev, the thermal activation energy, is about 10^{-2} of peak.

In the impurity region, the photoconductivity data taken at 77°K and at 22°K for a given sample differ only by a constant scale factor (3 for the data shown). This factor will reflect the changes in carrier lifetime and mobility with temperature. For the accuracy of our data, the shape of the photoconductive curves do not change in the 77°K–22°K interval. If a model employing a Coulombic potential is applicable, this is not surprising since the energy levels will only involve the dielectric constant as a variable. On this basis it would probably require at least a 2 percent change in dielectric constant or in turn approximately a 1 percent change in specific volume to observe a temperature effect.

Returning to Dunlap's model, one might expect that the response of *n*-type samples might be more complex. Here at least two ionization processes are possible. First, ionization of impurities which have but one excess electron and second, ionization of the second excess electron of impurities having two excess electrons. The form of the *n*-type response is perhaps suggestive of a superposition of two photoconductive spectra. In any event, the lowest energy threshold is compatible with the thermal activation energy reported by Dunlap. That is, a point about 10^{-3} down from the slowly changing section of the curve corresponds to an energy of 0.20 ev.

Conventional theories of quenching[3,4] (e.g. for the sulfides) center about the concept of a trapped hole which is freed by the quench light and subsequently recombines with a conduction electron to reduce the conductivity. It is presumed that the probability of recombination of a trapped hole and a free electron is small.

This theory has attractive possibilities for the present case. (The subsequent discussion will refer largely to measurements made on the 0.66-ev quenching effect.) In the dark the hole concentration and hence the concentration of trapped holes must be small in the *n*-type samples and little or no quenching is actually observed. The effect of background light is to increase the number of electrons in the conduction band and the number of free and trapped holes. If now the function of the quench light is to empty the hole traps,

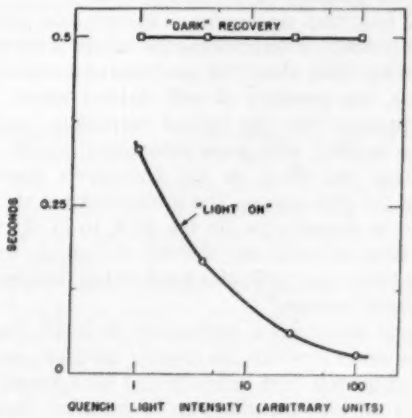

FIG. 11. Time constants for quenching at 22°K as a function of quench light intensity.

[4] A. Rose, RCA Rev. 12, 362 (1951).

than at the steady state there will be fewer uncombined electrons and holes and the background photocurrent will be quenched. (The argument could probably be reversed regarding holes and electrons except for the fact that quenching has been found for n-type samples having low resistance at 77°K which probably implies that any trap which was deep enough to be effective for this mechanism would already be saturated.)

If the procedure for dividing the effect of the quench light into normal and quench components is valid, then further evidence for the hypothesis is forthcoming. Presumably, the number of hole traps is a constant for a sample. It then seems reasonable that the quenching effect should saturate with increasing quenching intensity for a given background photocurrent. At a quench light level determined by the background, effectively all the traps will be emptied and no further changes in conductivity will occur. Similarly, a saturation should occur with respect to the background photoconductivity. That is, since the total number of hole traps is fixed, an upper limit should exist to the amount of conductivity that can be quenched; the fraction of the total conductivity that can be quenched should decrease with increasing conductivity. Also, at least at low background levels, the intensity of quench light necessary to produce a given fractional effect should increase more or less linearly with the background. Effects in qualitative agreement with these predictions are indicated by the data of Fig. 9.

We have no explanation to offer for the voltage characteristics of the quench curves. By the simple theory discussed above, the current-voltage characteristic should be linear. The fact that the characteristic for normal photocurrents is linear would seem to rule out the effects of contacts or inhomogeneities as the origin of the nonlinearity of the quenched photocurrents. The data suggest a decreasing efficiency for the quench process with increasing voltage.

If the hole trap mechanism is correct, one is faced with the problem of determining the nature of the traps. The striking thing about the spectrum of quenching is, of course, the presence of well defined bands. This would suggest that the optical transitions involved are "line to line" with some subsequent process communicating the effect to the continuous electronic states of the germanium. The insensitivity of the time constant to temperature (in the 22°K to 77°K range) would seem to rule out thermal activation from a discrete level to a continuous band as the time-limiting "subsequent process."

Without specifying a mechanism in detail, one can set up a model in which the discrete spectrum, and the effects of quench light intensity and background light intensity on time constant can be included. Suppose, for simplicity, that H holes produced by background radiation be trapped. Suppose further that the hole trap has a ground state occupied by h trapped holes and an excited state occupied by h' trapped holes. If the optical transition between them is allowed, then equilibrium between the states will be rapidly reached in any given radiation field. Suppose further that in the excited state the hole can ionize (perhaps by a multistep process) to a state p which can subsequently rapidly recombine with an electron. We assume that the process of ionization is the rate determining step.

Now, just after the quench light is turned on, since the $h' \leftrightarrows h$ equilibrium is rapidly established and the $h' \rightleftarrows p$ equilibrium is not, one can show that

$$h' = HB\rho/(2B\rho + A),$$

where B is an appropriate Einstein radiation constant, ρ the radiation density, and A a term which is the Einstein A plus a term describing any radiationless transition between h and h'. Similarly, under the aforementioned restrictions, the initial rate for the production of free holes will be of the form

$$dp/dt = k_{h' \to p}(h') + k_{\text{thermal}} + k_{\text{optical}} I - k_{p \to h'}(p) - p/\tau,$$

where the k's represent rate constants for the processes indicated in the subscripts, I is the intensity of background light and τ is the recombination lifetime. Substituting the expression for h' and assuming I large enough so that $p \sim k_{\text{opt}} I\tau$, in the absence of quench light, one then gets

$$\frac{dp}{dt} = k_{h' \to p}\left(\frac{HB\rho}{2B\rho + A}\right) - k_{\text{optical}} I\tau k_{p \to h'}.$$

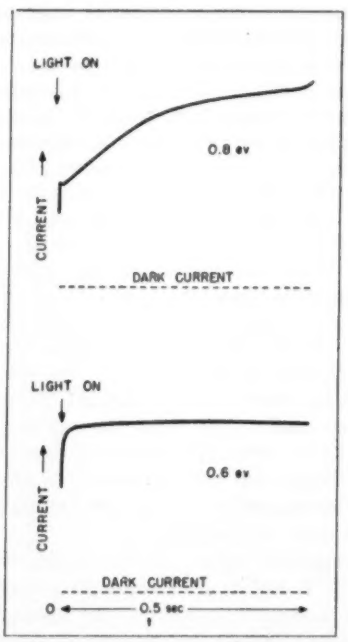

FIG. 12. Time dependence of response at 77°K of a sample showing quenching to 0.6-ev and 0.8-ev radiation. No background light present (from an oscillogram).

By the model, the rate of production of free holes is the rate of the quenching process. The above then indicates that the initial rate of quenching would show a linear increase with the intensity of quench light at low levels. Similarly, it should show a decrease with increased background level. Although no detailed confidence is placed in the model, it is of interest to note that the "light on" time constant curve of Fig. 11 resembles the predictions of the model at least at low quench intensities. Likewise, the "light on" time constant was observed to increase with background level. The time constant for recovery in the "dark" would be determined by the rate of the $p \rightarrow h'$ process, which should be independent of the previous quench light intensity.

A consequence of the hole trap model is that to the extent that the traps are involved, the fundamental photoconductivity should show long time constants identical with "dark" recovery times corresponding to the rate for trapping of free holes. On the other hand, for impurity photoconduction in the same sample, the long time effects need not be present. Figure 12 gives some confirmation of this view. It shows the "light on" response for a sample which exhibits quenching when illuminated with light in the fundamental region (0.8 ev) and when illuminated in the impurity region (0.6 ev), without any other radiation source present. Since the complete light on–light off cycle for the fundamental photoconduction does not show complete symmetry, presumably, some effects are involved which are even more complex than those discussed here. However, the major point seems to be established.

If a trapping process is involved in the quenching effect, the precise origin and nature of these traps remains unspecified. It is, for example, not completely established that the quenching explicitly involves the added gold. However, subsequent to the measurements described here, cases have been found where the interrelationships between quenching and trapping effects are more clearly brought out. The quenching spectra in these cases differ for that found for the gold and seem specific to the doping materials. The work will be presented in the near future.

It is a pleasure to acknowledge the assistance at various times of E. A. Taft. W. C. Dunlap, Jr., and L. Apker performed as sympathetic sounding boards for the author's ideas as well as offering many of their own. I am grateful to Professor Harvey Brooks for his criticism of this manuscript.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

High-Frequency Resistance of Thin Films*

R. BROUDY† AND H. LEVINSTEIN
Department of Physics, Syracuse University, Syracuse, New York
(Received December 15, 1953)

Measurements of high-frequency (hf) resistance of thin films have been performed. It is shown that the relatively small decrease of resistance with frequency occurring for uniform resistance is due to self-capacitance; the simple transmission line theory of this effect, originally presented by Howe, is incorrect. Additional decrease of hf resistance beyond that due to self-capacitance has been found to be due to the presence of nonuniform regions in the resistor and not to the presence of intercrystallite barriers in the frequency range considered. Properly interpreted, hf measurements can give information concerning the presence of gross inhomogeneities in films.

I. INTRODUCTION

WHEN a resistor is subjected to alternating current, the value of resistance decreases as the frequency increases. Several theories have been proposed to explain this behavior, but none has been satisfactory.

The extensive use of photoconductive cells in recent years and the search for a mechanism of photoconductivity in these cells has led to renewed activity in high-frequency (hf) measurements. The results of these measurements in cells have been used by many observers to substantiate their theories of photoconductive behavior; again there has been considerable disagreement concerning the significance of experimental results.

Hence it seems desirable to examine critically past data in order to obtain a mechanism that will explain observed results and to obtain information about the nature of thin conducting films.

The earliest reported investigations on high-frequency measurements of resistors were those of Boella[1] in Italy who observed additional high-frequency conductance

* This research was supported in part by the Wright Air Development Center. Based, in part, on a thesis submitted to the Graduate School of Syracuse University by R. Broudy in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

† Now at National Carbon Research Laboratories, Cleveland, Ohio.

[1] Mario Boella, Alta Frequenza **3**, No. 2 (April, 1934).

by using a modified resonant circuit. This phenomenon is often called "the Boella effect." Similar mesaurements were made in England by Sowerby and Marshall and reported by Puckle;[2] their measuring circuit was similar to that of Boella. A considerable diversity of results appeared between investigators and between different resistors tested.

Howe[3] originally developed a self-capacitance theory to explain the behavior of resistors at high frequencies in which he replaces the symmetrical electrical field lines along a resistor by capacitances and then folds it over to show the resemblance to a short-circuited transmission line. Further experimental and theoretical work on resistors was done to explain high-frequency resistance, in which capacitance to ground and of the end caps was considered.[4,5] Nevertheless, the theory was still rather unsatisfactory with many anomalous results. In cases of apparent agreement with theory, a combination of various factors had to be postulated, apparently without specific experimental verification.

Recently, high-frequency resistance investigations have received considerable impetus with their use in infrared photocells.[6–8] Chasmar,[6] and Humphrey, Lummis, and Scanlon,[8] have considered that much of the observed resistance decrease can be attributed to capacitive shunting across small crystallite barriers which are known to exist in thin films.

The present analysis will consider the resistance decrease expected because of barrier effects, the resistance decrease expected from simple self-capacitance effects, the resistance decrease expected from other effects, and information that may be obtained about the structure of thin films.

II. CRYSTALLITE-BARRIER EFFECTS

We will begin by developing a simple one-dimensional theory for capacitive shunting due to crystallite barriers only. Since the effect of electrodes (leading to self-capacitance, as will subsequently be explained) will at first not be considered, we call this an "infinite" film.

We set up the simple model for crystallites of a thin film shown in Fig. 1 and then consider that the film is made up of a succession of crystallites separated by high resistance barriers. Region 1 consists of a crystallite of conductivity σ_1 and length l. Region 2 consists of a barrier of conductivity σ_2 in series with region 1 and of length d. Subsequently, we will include the information: $\sigma_2 \ll \sigma_1$.

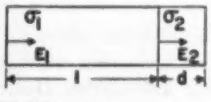

FIG. 1. One-dimensional crystallite barrier model.

[2] O. S. Puckle, Wireless Engr. **12**, 303 (1935).
[3] G. W. O. Howe, Wireless Engr. **12**, 291 (1935).
[4] L. Hartshorn, Wireless Engr. **15**, 363 (1938).
[5] G. W. O. Howe, Wireless Engr. **17**, 471 (1940).
[6] R. P. Chasmar, Nature **161**, 281 (1948).
[7] E. S. Rittner and F. Grace, Phys. Rev. **86**, 955 (1952).
[8] Humphrey, Lummis, and Scanlon, Phys. Rev. **90**, 111 (1953).

The boundary conditions which apply between regions 1 and 2 and which lead to a charge density at the boundary and hence to capacitive effects are

$$\epsilon_2 E_2 - \epsilon_1 E_1 = 4\pi q, \tag{1a}$$

$$\sigma_2 E_2 - \sigma_1 E_1 = -\partial q/\partial t, \tag{1b}$$

where E_2 = electric field strength in region 2, E_1 = electric field strength in region 1, ϵ_1 = dielectric constant of region 1, ϵ_2 = dielectric constant of region 2, σ_1 = conductivity of region 1, σ_2 = conductivity of region 2, and q = surface charge density at 1−2 boundary. Equation (1*a*) is a consequence of Gauss' theorem, and Eq. (1*b*), a consequence of the current continuity equation plus Ohm's law.

We now assume a harmonic solution to Eq. (1), i.e., all quantities are to be multiplied by $e^{i\omega t}$. For convenience, we use the same symbols, understanding that they now have phase and magnitude. Equations 1 become then

$$\epsilon_2 E_2 - \epsilon_1 E_1 = 4\pi q,$$
$$\sigma_2 E_2 - \sigma_1 E_1 = -i\omega q. \tag{2}$$

The potential across 1+2 is given by

$$V = E_1 l + E_2 d.$$

The current is given by

$$I = \sigma_1 E_1 A,$$

where A is the cross-sectional area. Then the admittance of crystallite+boundary is

$$Y = \sigma_1 E_1 A/(E_1 l + E_2 d). \tag{3}$$

We now eliminate q from equations (2) and obtain

$$E_2 = E_1(\sigma_1/\sigma_2)(\omega\tau_1 + i)/(\omega\tau_2 + i), \tag{4}$$

where

$$\tau_1 = \epsilon_1/4\pi\sigma_1, \quad \tau_2 = \epsilon_2/4\pi\sigma_2.$$

Substituting (4) in (3) leads to

$$G = \frac{\sigma_1 A}{l}\left[\frac{(\omega\tau_2)^2 + (d\sigma_1/l\sigma_2)\omega^2\tau_1\tau_2 + (1 + d\sigma_1/l\sigma_2)}{(\omega\tau_2 + [d\sigma_1/l\sigma_2]\omega\tau_1)^2 + (1 + d\sigma_1/l\sigma_2)^2}\right], \tag{5}$$

where G, the conductance, is the real part of Y. Replacing τ_1 and τ_2 by their equivalents, we obtain, after setting $\epsilon_1 = \epsilon_2 \equiv \epsilon$,

$$\sigma_f = \sigma_1\left[\frac{(\epsilon\omega/4\pi\sigma_2)^2(1 + d/l) + (1 + d\sigma_1/l\sigma_2)}{(\epsilon\omega/4\pi\sigma_2)^2(1 + d/l)^2 + (1 + d\sigma_1/l\sigma_2)^2}\right], \tag{6}$$

where σ_f = conductivity of crystallite+barrier. Now since the film is considered to be composed of these aggregates, σ_f is the film conductivity. Using the fact that $d/l \ll 1$ and manipulating (6), we have

$$\sigma_f/\sigma_0 = \frac{(\sigma_1/\sigma_0)^2 + (\epsilon\omega/4\pi\sigma_2)^2(\sigma_1/\sigma_0)}{(\sigma_1/\sigma_0)^2 + (\epsilon\omega/4\pi\sigma_2)^2}, \tag{7}$$

where, for $\omega = 0$, $\sigma_f = \sigma_0$.

It can be seen that the conductivity has the same general behavior that would be expected for a simple resistance in series with a parallel combination of resistance and capacitance. We are here interested in demonstrating only order of magnitude and so will not investigate the detailed behavior of (7).

Next, in order to determine the frequency range in which the conductivity increases significantly, we rewrite (7):

$$f=\omega/2\pi=(2/\epsilon)(d/l)(\sigma_1/\sigma_0)\sigma_0 \times\left[\frac{\sigma_f/\sigma_0-1}{\sigma_1/\sigma_0-\sigma_f/\sigma_0}\right]^{\frac{1}{2}}\left[\frac{1}{1-\sigma_0/\sigma_1}\right]. \quad (8)$$

Now consider an actual film for which $(\sigma_f/\sigma_0)_{\max} \equiv\sigma_1/\sigma_0\approx 20$. For such a film, the thickness is about 500A, $\epsilon\sim 4$ and the maximum crystallite size, l, about 2000A. d cannot be less than a few lattice spacings and we choose $d_{\min}$ about 10A. The frequency at which the film resistance drops to $\frac{1}{2}$ its value is given from (8):

$$f_{\frac{1}{2}}\sim 2.5(d/l)\sigma_0\sim(2.5)(0.005)(10^9)\sim 10^7.$$

Experimentally, for films which have this general sort of behavior, $f_{\frac{1}{2}}\sim 10^4$.

Since the approximations have been conservative and such as to make f small, we must conclude that the observed decrease of film resistance cannot be attributed to capacitive intercrystallite barrier effects. Simply stated, the time constant of crystallite and effective intercrystallite capacitance is not large enough to produce observable effects at the frequencies considered.

Other rough calculations were made for a case where the intercrystallite barrier is geometrically smaller than the crystallite; e.g., two spherically shaped crystallites in partial contact. These lead to the same qualitative $f_{\frac{1}{2}}$ results as above. This analysis thus shows that previous explanations of observed frequency variation by barrier shunting [6,8] are not valid.

Since, for the frequency range we are considering, these films are essentially uniform, we shall refer to them as "infinite uniform films." Next, we must consider practical films with connecting electrodes of essentially zero resistance and shall refer to them as "finite uniform films." In the next section, we will determine the capacitive shunting for this type of film.

III. FINITE UNIFORM FILMS

The discussion will concern a two-dimensional thin resistor and electrodes along the x axis:

$$\left.\begin{aligned} V&=+V_0, \quad l\leq x\leq\infty,\\ V&=-V_0, \quad -\infty\leq x\leq -l,\\ dV/dx+I(x)R&=0\\ dI/dx+\partial q(x)/\partial t&=0 \end{aligned}\right| \begin{aligned} &\\ &\\ -l\leq x\leq+l,&\\ \text{Resistor}&\end{aligned} \quad (9)$$

where $V(x)$=potential along x axis, $I(x)$=current along x axis, R=resistance per unit length, and $q(x)$=charge per unit area on the surface. Equation (9) is Ohm's law and the current continuity equation. Again, we assume a harmonic solution and use the same symbols for quantities having magnitude and phase; (9) becomes

$$dV/dx+I(x)R=0, \qquad dI/dx+i\omega q(x)=0. \quad (10)$$

Note that these equations are the same as those for a transmission line for which inductance and leakage conductance can be neglected, *if we could write* $q(x)=CV(x)$. Under these conditions the model originally proposed by Howe, in which he replaces electric field lines by capacitors and uses a transmission line calculation, would be valid. We claim, however, that the assumptions involved in setting $q(x)=CV(x)$ do not apply in the present case. The concept of a capacitance per unit length which appears in transmission line theory implicitly assumes that the capacitive coupling along a line length corresponding to the separation between conductors is negligible compared to that across the conductors. Actually the charge density along the resistor is given in the form

$$q(x)=\int_{-\infty}^{+\infty} k|x-x'|\,V(x')dx', \qquad \int_{-\infty}^{+\infty} q(x)dx=0. \quad (11)$$

From these relations and (10), one gets a complex integral equation not here treated.

The results for finite uniform films were obtained experimentally for material deposited in a controlled, uniform manner. Films were made of germanium, rhodium, and gold which were evaporated from a small molecular-beam oven, from a small bead on a large tungsten wire, and off the side of a cylindrical tungsten basket, respectively. By evaporating these materials in

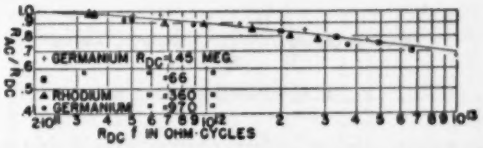

FIG. 2. Universal curve of R_{AC}/R_{DC} vs frequency times R_{DC} for uniform films. Points for some typical films.

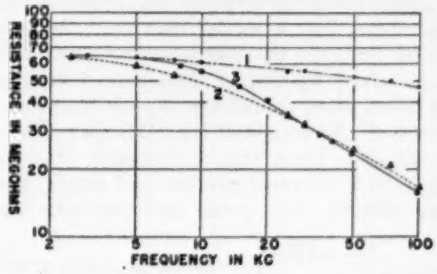

FIG. 3. Resistance vs frequency for: 1—a uniform film, 2—a typical nonuniform gold film, 3—Howe calculation.

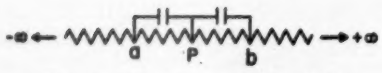

FIG. 4. An infinitely extending resistor.

this manner, films of uniform resistivity could always be obtained for germanium and rhodium and occasionally for gold.

We found that the data for all uniform films fit the curve of Fig. 2 which shows the ratio of ac to dc resistance *vs* frequency times R_{dc}. All points fall on this curve within experimental error. This curve is considerably higher than Howe's.

The data were taken with a General Radio 716-C capacitance bridge from 100 cps to 100 kc/sec, and a Boonton 160-A Q-meter from 50 kc/sec to 10Mc/sec.

The inadequacy of Howe's model is demonstrated in Fig. 3, which shows the resistance-frequency results expected from the Howe model and the experimental results for a typical film of the standard two-dimensional geometry. The capacitance per unit length to place in the transmission line calculation was determined by solving Eq. (10) with $\omega=0$ for $V(x,y)$. This involved the two-dimensional solution of Laplace's equation for boundary conditions (9). Then capacitance per unit length is given:

$$C(x)=-\frac{1}{4\pi}\left[\frac{\partial V(x,y)}{\partial y}\right]_{y=0}/V(x,0).$$

In order to show how one might obtain data that can apparently be explained by this model, Fig. 3 also includes the experimental results for a typical gold film of nonuniform resistivity. The significance of this curve will be discussed and explained in the next section.

IV. NONUNIFORMITIES

It has been observed that resistances of the same geometrical configuration exhibit different resistance-frequency behavior. In this section, we show that the additional ac conductivity beyond that of Fig. 2 is due to nonuniform resistivities of certain types.

We demonstrate first by a simple capacitance picture that any capacitive shunting is due to some form of nonuniformity. A resistor with constant resistance per unit length extending infinitely will develop no surface charge. From Fig. 4, we see that since the potential variation is linear, the charge at any point p due to points a and b equidistant from $p\equiv 0$. Pairing up all other points along the resistor, the net charge at $p\equiv 0$. This can also be seen from Eq. (11), since $V(x')$ is an odd function. From this, it becomes clear that a nonuniformity in resistance per unit length leads to charge density that would not otherwise appear.

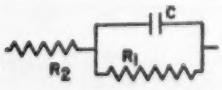

FIG. 5. Equivalent circuit for "abrupt" nonuniformity.

This analysis was substantiated by constructing films of known resistance variation. The materials and techniques used were the same as those for uniform films. The film of necessity must possess only non-uniformities that are purposely included.

First, an extremely sharp nonuniformity film was made with a high-resistance strip parallel to the electrodes at the center of the film. The strip resistance was 800 megohms and its width 2 mm, while the resistance and width of the remainder of the film on both sides of the strip were 200 000 ohms and 1.7 cm, respectively. The method of construction was to evaporate the low-resistance film, scrape off the 2-mm strip and re-evaporate till the film resistance, after aging, was 800 megohms. Since the nonuniformity is abrupt, it seems plausible to represent the film by the equivalent circuit of Fig. 5 where, in the present case, $R_1=800$ megohms (resistance of the strip), and $C=1.1$ $\mu\mu$f. The value of C was determined by measuring the difference in capacitance between a 2-mm gap and a

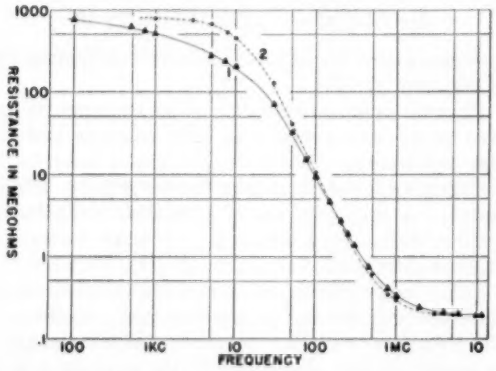

FIG. 6. Resistance *vs* frequency for "abrupt" non-uniformity film: (1) experimental, (2) calculated from circuit of Fig. 5.

1.7-cm gap. The parallel resistance for this circuit was calculated, and is shown along with the experimental data in Fig. 6; it can be seen that agreement is good considering the limitations of the simple model. Note that these curves are not fitted and the calculated curve is determined solely from measured values of R_1, R_2, and C.

Similar experiments and calculations were performed for various values of R_1, R_2, and C with similarly good agreement. Fig. 7 shows a family of curves calculated for the circuit of Fig. 5. These curves are similar to those of Humphrey, Lummis, and Scanlon, but are used in a different context and plotted differently.

Other films were made with a large number of similar gaps close to each other and some with randomly shaped gaps. In these cases, the slope of the resistance-frequency curve was not quite as large as that of Fig. 7, since the limitations of a lumped circuit model become more important. Characteristic behavior is such

that the hf resistance approaches that of the lower resistance sections in the manner of Fig. 7 as frequency increases. The behavior of these films can roughly be explained from a general consideration of Fig. 5. The time constant R_2C determines the higher frequency behavior ($R_1 \gg R_2$). This point was illustrated by evaporating films through different mesh screens. Then the hf resistance was measured; the dc resistance would have been about 1 megohm with gaps shorted. The measured ac resistance increased with increasing mesh and could not be detected above 60 mesh.

Actually, one would not expect to find only two resistance-per-unit-area regions in naturally irregular films; hence the resistance-frequency characteristic should continually drop off. A material for which this situation exists is gold, which is known to form irregular agglomerates. The nonuniformity was detected by painting in electrodes little by little and observing dc resistance. Figure 8 shows resistance-frequency curves for assorted gold films of identical geometry. The 40-megohm film was determined to be uniform by the

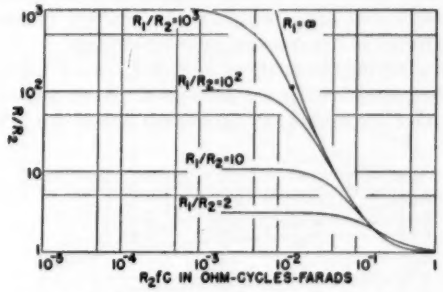

FIG. 7. Ratio of parallel resistance to dc resistance *vs* R_2fC for circuit of Fig. 5.

above described ohmeter method; its curve fits Fig. 2. Note again the nonuniform gold film curve included in Fig. 3

In addition, to show that nonuniformities need not be abrupt, films with smoothly varying resistance-per-unit-length were constructed by varying slide position relative to a point source. In all cases, the frequency decrease of resistance corresponded to the degree of variation in resistance-per-unit length and all curves fell somewhat below Fig. 2.

It was discovered, rather surprisingly, that resistance-frequency behavior does not depend strongly on the geometry. Finite uniform films were evaporated between

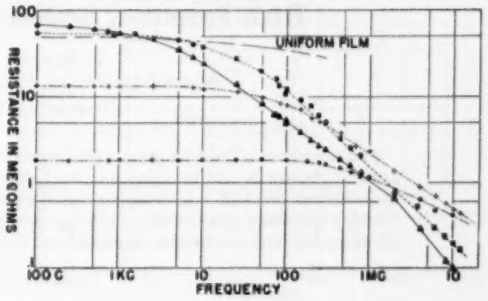

FIG. 8. Resistance *vs* frequency for assorted gold films.

various irregularly shaped electrodes; their resistance-frequency characteristics nevertheless followed closely that of Fig. 2.

CONCLUSIONS

The slight resistance decrease with frequency occurring for uniform resistances is due to simple self-capacitance. The previously evolved transmission line theory for this effect is invalid. The data of resistance *vs* frequency for uniform resistances fall on a universal curve where R_{ac}/R_{dc} is plotted against fR_{dc} (Fig. 2) for two-dimensional parallel electrodes. However, the effect is found to be geometrically insensitive.

The additional resistance decrease beyond that of the uniform case is due to the presence of several different resistance-per-unit-area materials in the film on a macroscopic scale. The previous theories of high-frequency conductivity due to capacitive shunting of intercrystallite barriers fail to explain observed results at the frequencies considered when subject to quantitative analysis.

In the past, many measurements of various properties of thin films have led to anomalous results. Some of these have been due undoubtedly to the presence of gross nonuniformities.

High-frequency resistance measurements on high-resistance films can demonstrate the presence of these types of inhomogeneities without disturbing the film itself and hence can be a valuable adjunct to many thin-film solid-state measurements. However, they can tell nothing of the crystallite or other small scale structure.

We are grateful to Professor M. Lax for many useful discussions.

High-Frequency Resistance of Photoconducting Films*

R. BROUDY† AND H. LEVINSTEIN
Department of Physics, Syracuse University, Syracuse, New York
(Received December 15, 1953)

Measurements of high-frequency (hf) resistance and photoconductivity have been performed on PbTe infrared photocells, and interpreted in the light of a recent analysis of hf resistance of thin films. Hf data show the presence of macroscopic nonuniformity in the photoconducting layers which decreases upon illumination. Since this technique will not detect the presence of inter crystallite barriers, their action in photoconductivity can neither be proved nor disproved by this method.

I. INTRODUCTION

IN recent years the results of high-frequency measurements on photocells have been used by various investigators to substantiate theories of photoconductive behavior. Chasmar[1] as well as Humphrey, Lummis, and Scanlon[2] have claimed that experimental results demonstrate the presence of intercrystallite barriers in photoconducting films. Chasmar[1] adduced that his results substantiated theories that barriers play an essential part in photoconductivity. Rittner and Grace,[3] however, stated that their own and Chasmar's results could be explained as being due to self-capacitance naturally occurring at a resistive film. Humphrey, Lummis, and Scanlon[2] obtained some data that was explained by self-capacitance alone; in other cases they used a combination of self-capacitance and barrier effects to explain observed effects. Since some cells did not exhibit barrier effects, they ruled out the presence of barriers as being essential to infrared photoconductivity.

We have recently found[4] that high-frequency measurements can tell nothing of crystallite or other similar small scale structure, and hence their results can neither prove nor disprove the action of intercrystallite barriers in photoconductivity. It was shown that additional high-frequency conductivity is due to the presence of macroscopic nonuniformities.

Then, having this analysis[4] available, we can interpret high-frequency resistance results of photoconducting cells. Our experimental data were taken for PbTe photocells, the results being similar to those previously obtained.

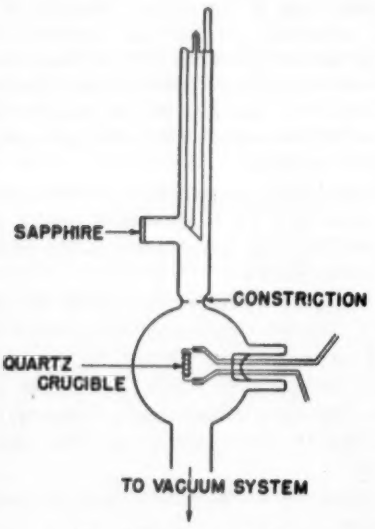

FIG. 1. Experimental cell.

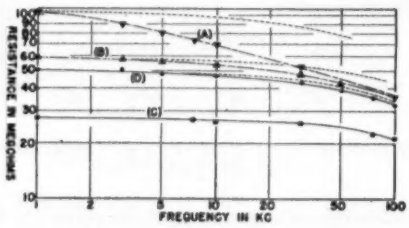

FIG. 2. Resistance vs frequency for special PbTe cell. (A) shielded; (B) exposed to room light; (C) illuminated; (D) after illumination, - - - finite uniform characteristics.

II. EXPERIMENTAL

In order to investigate the properties of the film material in a controlled manner, a cell was prepared by evaporating PbTe through a small hole in a quartz oven and then through a collimating slit onto the substrate with electrodes, such that material was deposited only across the parallel gap. Figure 1 shows a diagram of this cell.

Measurements were made with a General Radio 716-C capacitance bridge from 100 cps to 100 kc/sec and a Boonton Radio 160-A Q-meter from 50 kc/sec to 10 Mc/sec.

* This research was supported in part by the Wright Air Development Center. Based, in part, on a thesis submitted to the Graduate School of Syracuse University of R. Broudy in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

† Now at National Carbon Research Laboratories, Cleveland, Ohio.

[1] R. P. Chasmar, Nature **161**, 281 (1948).

[2] Humphrey, Lummis, and Scanlon, Phys. Rev. **90**, 111 (1953).

[3] E. S. Rittner and F. Grace, Phys. Rev. **86**, 955 (1952).

[4] R. Broudy and H. Levinstein, preceding paper [Phys. Rev. **93**, 285 (1953)].

Figure 2 shows resistance-frequency response for this cell under various conditions. The dc resistance with the cell shielded from room illumination was 110 megohms, and curve A was taken. The shield was removed and curve B resulted. Then the cell was moderately illuminated by flashlight and curve C taken. After removing the flashlight, the resistance rose to curve D. The dashed curve shows the characteristics of a finite-uniform film of the same dc resistance as each curve, taken from the previous analysis.[4] This curve falls exactly on top of C.

Utilizing the theory developed we can place the following interpretation on the data of Fig. 2: Curve A shows that the nonphotoactivated PbTe film has regions of high and low resistivity on a small but connected macroscopic scale. Upon being exposed to light (curve B) the conductivity of the regions of higher resistivity decreases and since this makes the resistance structure more uniform, the curve approaches more closely that of a finite-uniform film. If the film were

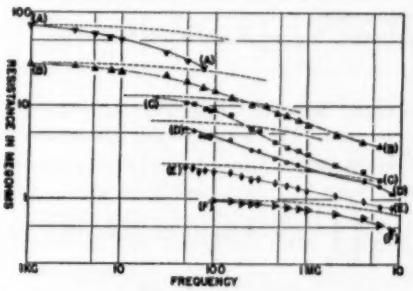

FIG. 3. Resistance *vs* frequency for a regular PbTe cell. (A) totally immersed in dark room; (B) totally immersed in lighted room; (C) exposed in dark room; (D) low additional illumination; (E) medium additional illumination; (F) high additional illumination. Simple dashed curves are finite-uniform characteristics.

equally photosensitive in all regions in the sense that the density of photocarriers is equal everywhere, one would expect curve B to be always below curve A, but approaching more closely a finite-uniform film since regions of higher resistivity would be decreased more percentage-wise. However, the fact that curve A falls on curve B at higher frequencies indicates that only the high-resistivity regions are affected by this type and intensity of illumination. Then, at a more intense illumination, both high- and low-resistivity sections of the film are affected, the high-resistivity sections now being reduced enough to make the film effectively uniform. Curve D shows the presence of a long-term photoeffect in both high- and low-resistivity sections. This datum and its interpretation correlates with sensitivity contours which show the presence of macroscopic regions of high and low sensitivity; these can be interpreted as regions of high and low resistance per unit area, respectively.

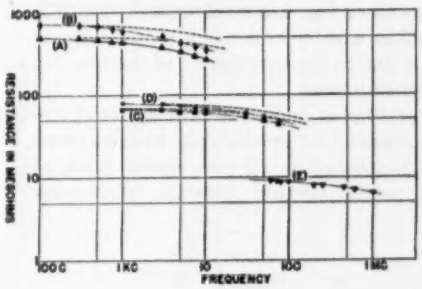

FIG. 4. Resistance *vs* frequency for a regular PbTe cell. (A) totally immersed in dark room; (B) in lighted room; (C) exposed in dark room; (D) in lighted room; (E) additional illumination. Simple dashed curves are finite-uniform characteristics.

Having demonstrated the presence of nonuniformities in the film itself, we can with some confidence take measurements on regular photocells if certain precautions are observed: during the processing of many PbTe and PbS cells considerable extraneous material, not connected to the electrodes, deposits inside the cell. Clearly this leads to spurious hf conductivity independent of that of the photoconducting film itself by being capacitively coupled; hence this type of cell cannot be tested by hf for film properties. Also, one is dubious about the nonuniformity introduced by the nonparallel geometry of the film since most cells do not have material deposited only across a parallel gap. However, in the latter case, we feel fairly secure in measurements on standard cells, since we have previously shown[4] that resistance-frequency behavior does not depend strongly on electrode-film configuration. In addition, considerable care must be taken to prevent condensation of moisture when testing cells which are cooled, since this is a source of hf conductivity.

Complete resistance frequency data were obtained for several PbTe cells. Figure 3 shows the results for a typical cell. Curves A and B were taken while the cell was totally immersed in liquid oxygen in a darkened and lighted room, respectively. Curve C was taken in a darkened room but not under total immersion conditions. Curves D, E, and F show the results for low, medium, and high illumination, respectively. As usual, dashed lines show finite-uniform curves for the same dc resistance. Again, nonuniformity decreases with increasing illumination. The same interpretation can be given to this behavior as that associated with Fig. 2, i.e., photocarriers cause greater percentage change in high-resistance film areas. Figure 4 shows similar data for another cell, clearly possessing a much more uniform film.

Several other cells, with negligible extraneous material, were similarly tested for resistance-photosensitivity-frequency. All results were similar to those of Fig. 3 and Fig. 4, falling somewhere between these two in departure from finite-uniform curves. The variation

for the cell of Fig. 4 is small enough that it is difficult to consider exactly how much of the extra hf conductivity is due to the irregularity of the film deposit outside the electrodes.

We attempted to correlate the degree of nonuniformity determined for several cells with pertinent characteristics: resistance, cell area, signal, noise, noise equivalent power, spectral response (photosensitivity *vs* wavelength). It was found that there was no correlation with any of these properties except spectral response; in this case, the degree of irregularity in the response *vs* wavelength curve varied with the degree of nonuniformity of the film. It seems reasonable to explain this by considering that the photosensitive films possess several sensitive regions, each independently contributing to the total photosensitivity and spectral response.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Electric Polarization of Charged Particles in Square Potential Wells

JULIAN H. GIBBS
General Electric Company, Schenectady, New York
(Received August 11, 1953; revised manuscript received January 15, 1954)

The Schrödinger equation is considered for the case of a (charged) particle which is rigidly restricted to a finite region within which it is subject to a linear (electric) potential. Expressions are obtained for the energy and polarization of the quantum levels in the case where the field is small. Exact values for these properties in a stronger field are calculated for a few representative cases.

The properties of a canonical assembly of such systems are treated. In particular, the polarization of such an assembly is determined as a function of temperature.

I. INTRODUCTION

IN connection with a physical problem which will be discussed in a future paper, it was of interest to determine the effect of potential type on the polarizability behavior of an assembly of charged particles in unidimensional potential wells. Of the potentials considered, one, which yielded a tractable quantum-mechanical problem, was the square well with infinitely high walls. In this system the electric field appears as a uniform slope in the floor of the otherwise rectangular potential box.

The treatment of this general quantum-mechanical system and the assembly of such systems has been removed from the context of the particular physical application which motivated the work and is presented independently in this article.

In Sec. II, the Schrödinger equation for this system is solved by treating the electric field as a perturbation. Expressions are obtained for the energy and polarization of the various states of the system.

In Sec. III, the polarizability of an assembly of such systems is obtained as a function of temperature.

In the final section, the corrections which must be applied to these results for the lowest levels in the case of strong field are calculated numerically for a few cases.

II. ENERGY AND POLARIZATION OF THE LEVELS IN SMALL FIELDS

In this case we may write, for the energy of the nth level, $\epsilon_n \simeq \epsilon_n^0 + \epsilon_n''$, where $\epsilon_n^0 = n^2h^2/8Ma^2 = Kn^2$ in which M is the mass of the particle and a the width of the potential box; the first-order correction to the energy is zero. The second-order correction to the energy ϵ_n'' is given by

$$\epsilon_n'' = \frac{\sum_m{}' \left[\int_0^a (2/a) zFx \sin(m\pi x/a) \sin(n\pi x/a) dx\right]^2}{K(n^2-m^2)},$$

for a particle of charge z in the presence of a field F acting along the one dimension x.

The matrix elements may be integrated by parts to give

$$\epsilon_n'' = -(256Ma^4z^2F^2n^2/\pi^4h^2) \sum_{\substack{m=-\infty \\ m\pm n \text{ odd}}}^{+\infty} m^2/(m^2-n^2)^5.$$

Expanding the ratio under the summation sign in partial fractions, and interchanging the order of the summations, we have

$$\epsilon_n'' = (-256MA^4z^2F^2/\pi^4h^2) \sum_{\beta=1}^{5} \left\{ n^{\beta-6} C_\beta \left[\sum_{\substack{m=-\infty \\ m-n \text{ odd}}}^{+\infty} 1/(m-n)^\beta + (-1)^\beta \sum_{\substack{m=-\infty \\ m+n \text{ odd}}}^{+\infty} 1/(m+n)^\beta \right] \right\}.$$

Inasmuch as each of the sums inside the brackets is taken over the reciprocals of all odd numbers, raised to the power β, the whole quantity in brackets is zero for $\beta=1, 3, 5$. It is equal to $\pi^2/2$ for $\beta=2$ and $\pi^4/24$ for $\beta=4$. Evaluation of the partial fraction coefficients, C_β, gives $C_2=5/256$, $C_4=-1/64$ so that we have, for the

perturbation energy,

$$\epsilon_n'' = (-Ma^4z^2/6\pi^2h^2)[(15/n^4)-(\pi^2/n^2)]F^2$$
$$= (A/2)[(15/n^4)-(\pi^2/n^2)]F^2,$$

or, for the polarization of the nth energy level,

$$P_n = -\partial\epsilon_n/\partial F = (A)[(15/n^4)-(\pi^2/n^2)]F = \alpha_n F,$$

α_n denoting the polarizability of this nth level.

The negative polarizabilities exhibited by all levels above the ground state are subject to a simple classical interpretation. A classical particle whose energy is great enough to permit it to go anywhere in the region $0<x<a$ will actually spend more time in a neighborhood of high potential energy, where it must move slowly, than in a more rapidly traversable one of lower potential energy.

III. AVERAGE POLARIZABILITY IN THE ASSEMBLY

The average polarizability in a canonical assembly of such systems is given by

$$\alpha = A\frac{15\sum_{n=1}^{\infty} n^{-4}\exp(-Kn^2)-\pi^2\sum_{n=1}^{\infty} n^{-2}\exp(-Kn^2)}{\sum_{n=1}^{\infty}\exp(-Kn^2)},$$

the approximation of ϵ_n by ϵ_n^0 in the exponentials being justifiable in this small field case. The sums appearing in this expression may be evaluated by means of their connection with the tabulated theta function, $\theta_3(v) = \sum_{n=-\infty}^{+\infty} q^{n^2}e^{inv/2}$. That is, if we take $v=0$, $q=e^{-K}$, we have

$$\sum_{n=1}^{\infty}\exp(-Kn^2) = \tfrac{1}{2}[\theta_3(0)-1].$$

Furthermore,

$$\sum_{n=1} n^{-2}\exp(-Kn^2) = \tfrac{1}{2}\int_K^{\infty}[\theta_3(0)-1]d\lambda,$$

$$\sum_{n=1}^{\infty} n^{-4}\exp(-Kn^2) = \tfrac{1}{2}\int_K^{\infty}(\lambda-K)[\theta_3(0)-1]d\lambda,$$

in which we now understand $q=e^{-\lambda}$.

Therefore, by carrying out the indicated integrations between K and infinity for various K, we may obtain α as a function of K (or $1/T$) if a remains constant. This has been done, and, in Fig. 1, the quantity α/A is plotted *vs* K. It will be seen that, for small K, α/A varies as K.[1] This is the classical limiting law for this system.[2]

[1] This may be demonstrated analytically by Euler-MacLaurin expansion of $\sum_{n=1}^{\infty}\exp(-Kn^2)$ followed by the indicated integrations to obtain expressions for the other sums involved in α.

[2] The classical law for the polarization is

$$P = (Nza/2)[\coth(zaF/2kT)-(2kT/zaF)]$$

which reduces to $P=N(za/2)^2F/3kT = N\mu_{max}^2F/3kT$ for high T.

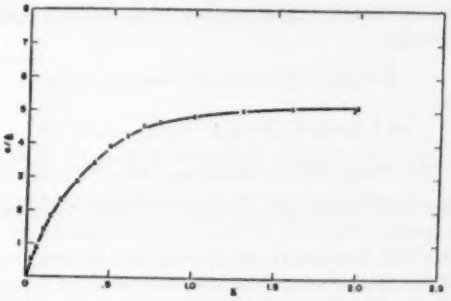

FIG. 1. α/A *versus* K.

At higher K (lower T) quantum deviations set in, and the polarizability no longer increases as rapidly with decreasing T. The asymptotic value which is approached at very low T is, of course, simply the average polarizability of N such systems all in their lowest-energy levels.

IV. ENERGY LEVELS FOR THE CASE OF STRONG FIELDS

In the case of strong fields we can no longer use the perturbation approximation. Instead, however, we may obtain numerical results by application of the box boundary conditions to the exact solution to the Schrödinger equation for a particle in a constant field of force. This solution is known[3] and, with the zero of

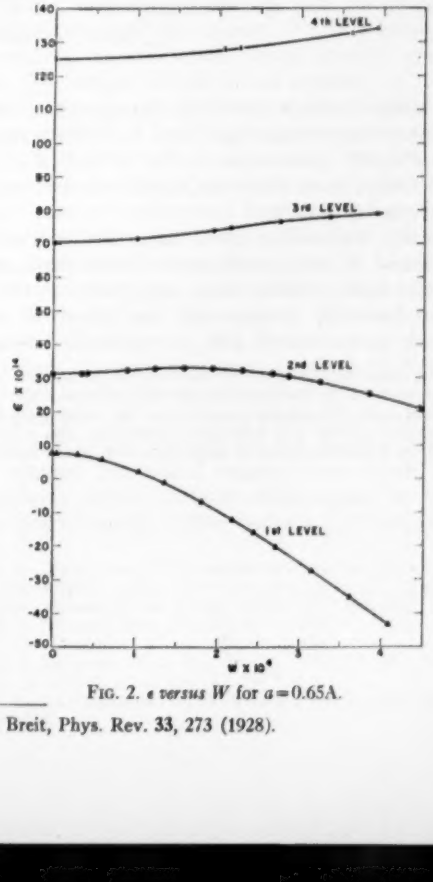

FIG. 2. ϵ *versus* W for $a=0.65$A.

[3] G. Breit, Phys. Rev. **33**, 273 (1928).

potential energy taken at the point $x=a/2$, may be written as

$$\psi=(By^{\frac{1}{2}}/c)[\sin\phi J_{\frac{1}{3}}(y)+\cos\phi J_{-\frac{1}{3}}(y)],$$

where

$$y=\tfrac{2}{3}(8\pi^2MW/h^2)^{\frac{1}{2}}[x-(a/2)-(\epsilon/W)]^{\frac{3}{2}},$$
$$y^{\frac{1}{2}}/c=[x-(a/2)-(\epsilon/W)],\quad\text{and}\quad W=-zF.$$

$B\sin\phi$ and $B\cos\phi$ are the two arbitrary constants of this general solution; B may be determined by normalization. In the present case, ϕ and discrete values, ϵ_n of the energy ϵ are determined by the box boundary conditions,[4] $\psi(y)=0$ at $x=0, a$.

It should be noted that the transition from a wave function of oscillatory character to one which falls monotonically in the region of barrier penetration is characterized by a change from real to imaginary argument, y_n, of the Bessel functions at the value of x such that $\epsilon_n=[Wx-(a/2)]$. Higher energy levels for which $\epsilon_n>-|zFa/2|$, the largest finite value of the potential energy, do not, of course, exhibit this change, y_n remaining real throughout.

The numerical procedure followed was to choose various values of ϕ, find the roots, y_n, of $\psi(y_n)=0$ for each value of ϕ, by inspection of tables of the appropriate Bessel functions, and then to determine the values of ϵ_n and W which fit the spacing of these roots to the boundary conditions of the problem for a given value of a by means of the pair of equations,

$$y_1^{\frac{1}{2}}=(32\pi^2MW/9h^2)^{\frac{1}{2}}[(-a/2)-(\epsilon_n/W)],$$
$$y_{n+1}^{\frac{1}{2}}=(32\pi^2MW/9h^2)^{\frac{1}{2}}[(+a/2)-(\epsilon_n/W)].$$

Here y_1 is the imaginary root, if one exists; otherwise it is the real root of smallest absolute magnitude. The other roots are ordered according to increasing n as their absolute magnitudes increase.

The results for a proton with $a=0.65$A are shown in Fig. 2 in which the values of the various ϵ_n are plotted against the corresponding values of W. The slope of the curve for each level, with sign changed, is a measure of the polarization of that level. It is interesting to note that the second level behaves like the upper levels in having a negative polarization for small W but reverses itself for larger W and exhibits a positive polarization like the ground level in this region.

In the calculation of the polarization of the assembly for the strong-field case, it is still possible to use the results of perturbation theory for the upper levels. The contribution of the r lowest levels to the assembly polarization may be corrected by use of the following formula:

$$P=N\frac{\left\{\sum_{n=1}^{r}[(-\partial\epsilon_n/\partial F)\exp(-\epsilon_n/kT)+(\partial\epsilon_n''/\partial F)\exp(-\epsilon_n^0/kT)]+\sum_{n=1}^{\infty}(-\partial\epsilon_n''/\partial F)\exp(-\epsilon_n^0/kT)\right\}}{\left\{\sum_{n=1}^{r}[\exp(-\epsilon_n/kT)-\exp(-\epsilon_n^0/kT)]+\sum_{n=1}^{\infty}\exp(-\epsilon_n^0/kT)\right\}}.$$

A similar formula, in which the quantities $\partial\epsilon_n/\partial F$ (slopes of the curves in Fig. 2) and $\partial\epsilon_n''/\partial F$ are replaced with $\partial^2\epsilon_n/\partial F^2$ (curvatures in Fig. 2) and $\partial^2\epsilon_n''/\partial F^2$, respectively, gives the polarizability of the assembly in this strong-field case.

Finally, it should be noted that if the field, or W, is considered as being composed of two parts, one a constant (and possibly large) contributor to the individual potential systems and the other the small, variable applied electric field, we have here the solution to the problem of the electric polarizability of an assembly of particles in square wells with intrinsically sloped floors, as well as the solution for the polarizability of the assembly of simple square wells. It is actually this solution which was particularly desired for the physical application mentioned in the Introduction. An abstract of this application, a theory of ferroelectricity in Rochelle salt, has already appeared.[5]

The author wishes to express his thanks to Dr. J. Rosenbloom of the University of Minnesota and to Dr. E. W. Hart, Dr. B. H. Zimm, and Dr. J. K. Bragg of the General Electric Research Laboratory for many helpful discussions and suggestions.

[4] The quantum rule obtained above for the small field case by perturbation of the solutions for the box potential could also be obtained from the general solution for the linear potential by application of the box boundary conditions, the asymptotic forms for Bessel functions of large argument (small field) being used.

[5] J. H. Gibbs, Phys. Rev. **91**, 447 (1953).

Thermionic and Surface Properties of Tungsten Crystals*

GEORGE F. SMITH†
California Institute of Technology, Pasadena, California
(Received November 9, 1953)

A plateau-like surface structure was found on single crystals grown in tungsten wire. Plateaus were observed on the surface normal to the (110), (112), and (001) directions. A "shingle" structure was observed superimposed upon the plateaus. A wire subjected to only mild heat treatment showed both types of structure, indicating that evaporation cannot be the only mechanism involved. All wires were heated by ac exclusively to avoid the "dc etch."

New values of the thermionic constants for different crystal directions were measured on a crystal having a minimum of shingling. Measurements were made in the tube previously used by Nichols, modified to eliminate an anomalous effect. The modification did not appreciably alter the emission constants for the (111), (112), (116), and (001) directions, which constants are in reasonable agreement with Nichols' values. A second anomalous effect contributed to currents measured in the low emission (110) direction, rendering normally obtained (110) values erroneous (including those of Nichols). Although the true (110) constants could not be obtained, an upper limit for the emission is characterized by $\varphi^{**}=4.72$, $A^{**}=9.7$ and two estimates of about 5.3 v are deduced for φ^{**}. There was no evidence in general for patchiness of the single crystal surfaces investigated; however, the usual tests cannot conclusively deny the existence of patchiness.

I. INTRODUCTION

THE technique of producing crystals of centimeter length in tungsten wire is well known. As a result of about 1 percent of "dope" consisting of Na_2O, K_2O, $CaCl_2$, Al_2O_3, and SiO_2 added before the sintering process,[1] the "nonsag" type of tungsten wire recrystallizes into crystals occupying the entire cross section of the wire. Robinson[2] has shown that under a favorable vacuum heat treatment schedule one can produce crystals several centimeters long in prewar nonsag tungsten wire. Chemical analyses of doped tungsten wire after the usual heat treatment associated with thermionic emission measurements show about the same very small relative impurity content as in undoped wire.[3] This fact along with thermionic emission stability through all temperature cycles after sufficient heat treatment, gives assurance that emission studies of doped wire are actually characteristic of pure tungsten.[3] As a result of the working given the wire during the drawing process, the crystals grown in doped wire are normally oriented with a face diagonal very nearly parallel to the wire axis. This orientation is particularly favorable to emission studies since, if the wire is smooth and round, the placing of the face diagonal along the wire axis causes directions emerging normal to the surface to include all directions with Miller indices of the form (*hhk*). Martin's work[4] with thermionic emission from a single crystal sphere of tungsten showed that the directions of principal interest, namely the maxima and minima of emission, are all of the form (*hhk*).[5]

Two kinds of surface structure have been observed on single crystals of tungsten heated in vacuum: (a) a so-called "dc etch," and (b) a "shingle" structure. The former effect was first reported by Johnson,[6] and has been studied by Schmidt[7] and by D. Langmuir.[8] It consists of a step-like structure found on wires subjected to prolonged heating with direct current; it also occurs near the supports of a wire heated with alternating current, where strong thermal gradients exist. The shingle structure was reported by Nichols[9] and was observed in the (110) crystal direction. Both structure effects were to be found on the crystal from which Nichols[9] obtained his emission constants as a function of crystallographic direction. Moreover, prior to the present investigation it has appeared plausible to explain the emission data for the (110) direction as being characteristic of a patch surface,[9,10] contributory evidence of the patchy nature deriving from the existence of the shingle structure on that surface. The principal objective of this research was to make thermionic measurements from a crystal (a) free from the dc etch and (b) with minimized shingle structure. Freedom from dc etch could readily be achieved by using only ac for filament heating. Indeed, Johnson[6] has reported that helical tungsten wires heated by ac in vacuum develop surfaces which appear to be free from any structure visible under a light microscope,

* This work was done with a facility created with the aid of a Research Corporation grant and was supported in part by the U. S. Office of Naval Research. It has been reported in greater detail in Final Report N6-onr-24433, June, 1952 (unpublished). Part of the work was done in partial fulfillment of the requirements for the Ph.D. degree at the California Institute of Technology.

† Now at Hughes Research and Development Laboratories, Culver City, California.

[1] C. J. Smithells, *Tungsten* (D. Van Nostrand Company, Inc., New York, 1936), second edition.

[2] C. S. Robinson, J. Appl. Phys. **13**, 647 (1942).

[3] M. H. Nichols, Phys. Rev. **78**, 158 (1950).

[4] S. T. Martin, Phys. Rev. **56**, 947 (1939).

[5] J. A. Becker, in private communication, has indicated that his field emission studies show that there may be a small field emission minimum in the vicinity of the (210) direction.

[6] R. P. Johnson, Phys. Rev. **54**, 459 (1938).

[7] R. W. Schmidt, Z. Physik **120**, 69 (1943).

[8] D. B. Langmuir, Phys. Rev. **89**, 911 (1953); **91**, 447 (1953).

[9] M. H. Nichols, Phys. Rev. **57**, 297 (1940).

[10] Conyers Herring and M. H. Nichols, Revs. Modern Phys. **21**, 185 (1949).

even after heating sufficient to evaporate away 17 percent of the wire. The shingle structure, however, has been found to occur on wires heated only with ac. Nevertheless, it was believed that this type of structure could be eliminated or at least greatly reduced by a careful examination and selection of the crystal to be used.

II. GROWTH AND SURFACE STRUCTURE OF SINGLE CRYSTALS OF TUNGSTEN

With this purpose in mind, a program of single-crystal production was instituted. Wires to be recrystallized were first polished[11] to remove die-marks and to produce a uniform circular cross section. The recrystallization process itself was carried out in what is commonly called an electron projection tube[12,9] after the method of Robinson.[2]

Six 14-in. lengths of 0.005-in. tungsten wire were recrystallized. Of the six, four were of G.E. No. 218 wire, dated 1939, each of which produced crystals several centimeters in length. The remaining two were of Callite No. 200*H*, dated 1940, neither of which produced crystals of more than millimeter length. One G.E. wire was recrystallized at 1980°K, producing a single, but relatively imperfect, crystal filling the 14-in. wire. The 2100°K runs each produced several crystals averaging three or four centimeters in length. In each case the orientation, length, and degree of perfection of single crystals produced were carefully noted from the emission pattern as observed in the projection tube. Cathetometer measurements were made to fix the position of each crystal section that showed an emission pattern free from flaws. Then the wire was removed from the projection tube, and each selected section was examined for structure visible under a light microscope of 100× magnification.

In each case it was possible to resolve shingle structure associated with the (110) direction, even in the case of wires subjected to only mild heat treatment. The 1980°K wire, given a total heat treatment of 22 hr at 1980°K plus 20 min at 2280°K, should have suffered an average diameter reduction of about 3×10^{-8} cm.[13] On this wire not only was the shingle structure observed on the two surfaces normal to (110) directions, but the shingles were superimposed upon what appeared to be a plateau, or flat surface, on the crystal. Even where the shingling was particularly faint, one could make out the plateau edges, separated by a few degrees of azimuth. Before prolonged heat treatment all recrystallized wires showed both these effects, as well as grain boundaries and occasional surface crystal "lakes."

The shingle structure observed in recrystallized wires before prolonged heat treatment varied both in form and visibility. The shingling is possibly a result of the failure of the crystal face diagonal to be aligned *precisely* along the wire axis. The orientation of the apexes of the shingles then indicates the direction of "dip" of the true crystal planes with respect to the wire surface. If a true cylinder were cut from a perfect

FIG. 1. Projection tube pattern of wire containing single crystal section used for quantitative thermionic measurements. The white marks set off the single crystal section; the dark marks are crayon lines on the tube for identification purposes.

[11] Johnson, White, and Nelson, Rev. Sci. Instr. **9**, 253 (1938).
[12] R. P. Johnson and W. Schockley, Phys. Rev. **49**, 436 (1936).
[13] H. A. Jones and I. Langmuir, Gen. Elec. Rev. **30**, 310, 354, 408 (1927).

crystal, one would expect to find opposite directions of dip and opposite orientation of shingle apexes on the opposite (110) surfaces. This was usually but not always found to be the case for the crystals examined. The few instances noted with the same apex orientation on both sides of the wire could be due to the presence of an undetected longitudinal grain boundary between the two sides of the wire; variation in wire diameter along the length of the crystal could also produce this effect. Longitudinal grain boundaries are common in in G.E. No. 218 wire, and would be difficult to observe on a wire given only mild heat treatment, especially if the orientations of the two crystals were not appreciably different; no examination of this phenomenon was made after prolonged heat treatment. In a few cases a (110) surface was found with a shingle-free region having oppositely oriented shingles at either end. A slight bend in the wire before recrystallization could produce this effect; Johnson[6] has reported that a single crystal grown in a tungsten helix maintains a fixed orientation in space as though the helix were carved out of a large crystal. In some isolated cases surfaces were found with little or no shingling in evidence over several millimeters along the axis of the wire. Both in obtaining a shingle-free (110) surface and in having precise (*hhk*) directions exhibited normal to the wire axis, such a section would show promise for quantitative thermionic measurements. At this point one of the four G.E. No. 218 wires was selected for further investigation. The recrystallization schedule of the wire is given in Table I, A; the projection tube pattern after recrystallization is shown in Fig. 1.

TABLE I. History of single crystal wire. Temperatures below 2100°K during taking of thermionic data are not listed.

Heat treatment	Subsequent observation
A. 2100°K, 15.5 hr; 2410°K, 10 min; 2610°K, 15.5 min; 2820°K, 0.5 min.	Plateaus in (110) directions, shingling.
B. 2610°K, 240 min; 2720°K, 30 min; 2820°K, 10 min; 2915°K, 2 min.	Little change in structure. Repolish.
C. 2760°K, 340 min; 2990°K, 10 min.	Similar structure restored.
D. 2700°K, 165 min; 2750°K, 51 min; 2830°K, 5 min.	Discover (001), (112) plateaus optically.
E. 2575°K, 32 min; 2775°K, 34 min.	Pre-end-guard thermionic data.
F. 2785°K, 38.5 min.	Final thermionic data.
G. 3030°K, 1 min; 2800°K, 5 sec; 1230°K, 92 hours.	Photograph plateaus, see Fig. 2.

To determine the effect of moderate heat treatment upon the size and shape of surface structure, the wire was subjected to the temperature schedule of Table I, B. Although this treatment left the surface in general almost entirely smooth and free from markings, the size and appearance of the plateaus and shingle structure were not materially altered. The recrystallized wire was then repolished (the diameter was reduced 1 percent) in order to determine whether the structure had been produced during recrystallization, not to return if once removed. After the relatively severe heat treatment of Table I, C (sufficient to evaporate about 2 percent of the diameter)[14] the plateau and shingle structure were observed to be restored to essentially the same condition as observed before repolishing. From the wire in this condition a single crystal section was selected for the quantitative thermionic measurements. The section chosen showed a minimum of shingling and lake structure together with a flawless qualitative emission pattern. This section is marked on Fig. 1. The further heat treatment listed in Table I was dictated by outgassing procedures. Very faint plateaus were also discovered optically in the (001) and (112) directions after a preliminary run in the thermionic measuring tube. At this time the wire had suffered less than 0.5 percent additional diameter reduction by evaporation due to the temperature schedule of Table I, D, making the total reduction about 2.5 percent from its value at the conclusion of repolishing.

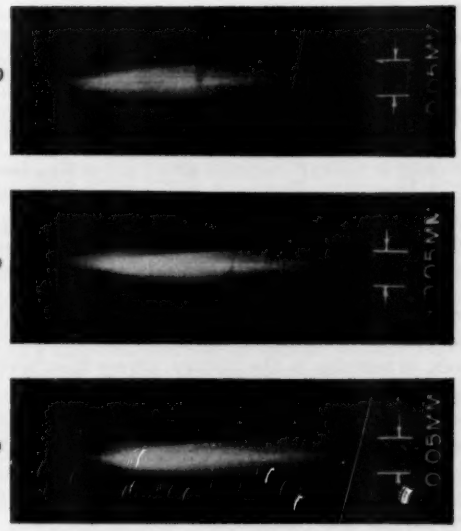

FIG. 2. Photomicrographs of single crystal showing plateau structure. (a) 180° (110) plateau; (b) 235° (112) plateau; (c) 90° (001) plateau.

At the conclusion of the final thermionic emission measurements (see Sec. V), subsequent to the heat treatment of Table I, EFG, the photomicrographs of Fig. 2 were taken of the surface structure on the chosen crystal section. Illumination at grazing incidence, obtained by adjusting the direct-dark field slide of the Zeiss Epi condenser microscope midway between the two usual positions, was found necessary to provide sufficient contrast to make out the smaller structure at all. The dimensions marked on the photographs and plateau widths listed in Table II were obtained by photographing an Eastman stage calibrated scale; the

[14] As measured by micrometer. From the Jones-Langmuir tables, reference 13, one computes a reduction three times as large. In general the evaporation computed by the Jones-Langmuir tables was greater than that observed by diameter measurements.

TABLE II. Plateau widths after heat treatment of Table I.

Direction of plateau	Width from photomicrograph	Width from emission plot	Orientation of shingling
0° (110)	11°	...	Left, faint
180° (110)	11°	Invalid (see Sec. IV)	Right, faint
55° (112)	7°	6.8°	None resolved
125° (112)	7°	6.3°	Right, very faint
235° (112)	7°	6.3°	None resolved
90° (001)	7°	7.8°	Right, faint

error in these measurements is probably less than 20 percent. The photomicrograph of the 90° (001) surface shows, in addition to the plateau, a very faint periodic diagonal scoring which has the pitch of the polishing machine screw. Although the average wire diameter at this point had been reduced about 3.6 percent by evaporation from the value after repolishing, it appears that faint markings in the vicinity of the (001) surface survived the heat treatment.

Additional evidence of the existence of (110), (112), and (001) plateaus is to be found in polar plots of thermionic emission *versus* azimuth around the wire. Thermionic plateaus were discovered during the first sufficiently detailed preliminary thermionic measurements. These measurements were taken just before the conclusion of Table I, D, i.e., just prior to optical observation of the (112) and (001) plateaus. The final data polar plots of Fig. 8 show "flats" of thermionic emission in the (110), (112), and (001) directions; plateau widths can be had from Fig. 8 by adding 2.3°, the resolution of the emission measuring tube, to the widths of the emission flats. The values listed in Table II are taken from the 1810°K curve; they are seen to be in good agreement with those obtained from the photomicrographs. Studies of field emission from tungsten points of the order of 10^{-5} cm radius have shown emission flats in the same directions.[15–17]

At the same time that the photomicrographs were made, the diameter was measured by a simple optical wedge interferometer. The results are plotted as a function of azimuthal angle in Fig. 3. A definite upward departure from the average diameter is observed in the (001) direction and there may be a small upward departure in the (110) direction, but no minima are seen at the centers of these maxima. Note however that the deviation in diameter expected over an 11° plateau is 0.5 percent, or about two fringes of mercury green light, while that expected over a 7° plateau is one fringe. Moreover, the crystal section measured was located at the center of a 3.5-in. length of wire; although care was exercised in lowering the optical flat onto the wire, some slight twisting of the wire might have occurred if the flats came to rest in such a way as to impose a torque on the wire.

Just prior to removal from the emission measuring tube, the crystal was subjected to the temperature schedule of Table I, G, in an attempt to find some variation in plateau size as a function of temperature. Becker[16] has shown that on tungsten field emission points of about 3×10^{-5} cm radius, the steady state linear dimension of the (001) and (112) planes increases by a factor of almost two as the temperature is lowered from 2600° to 1200°K. He also found that the time necessary to reach a steady state increased as the temperature was lowered, a few seconds or less being required at 2600°K, while at 1200°K several minutes are needed. Since our tungsten wire is much larger than Becker's field emission point, the times used were much longer than his steady-state times. The 3030° and 1230°K heating periods were each preceded and followed by a rapid emission measurement, at 1560°K, of the widths of the 55° (112) and 90° (001) plateaus; the reading of the two widths required approximately 10 min. The three sets of width measurements agreed with one another to 0.1°, and indicate no experimentally observable change in plateau size under the stated limited temperature schedule.

III. DISCUSSION OF SURFACE STRUCTURE

The kind of surface structure observed on tungsten has been found by Nichols[18] on tantalum single crystals heated in vacuum, although no plateaus were observed on tantalum normal to the (112) directions. He has also found plateaus on molybdenum crystals. The formation of plateaus on tungsten, tantalum, and molybdenum must be due to one or more of three processes: (1) differential evaporation, (2) volume diffusion, and (3) surface diffusion. Herring[19] proposes that volume diffusion motivated by surface tension is mainly responsible for the plateau growth; he has developed, under certain reasonable assumptions, a theory that predicts remarkably well the size of tungsten and tantalum plateaus produced under different temperature schedules. Clearly, evaporation has no

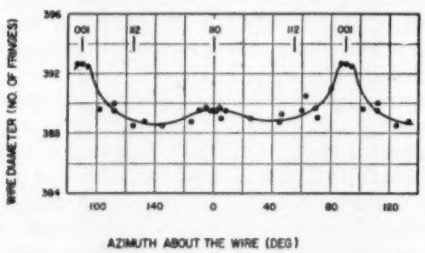

FIG. 3. Wire diameter as a function of angle around the wire. A simple wedge interferometer was used with mercury green light.

[15] E. W. Müller, Z. Physik **120**, 261, 270 (1943).
[16] J. A. Becker, Bell System Tech. J. **30**, 907 (1951); also by private communication.
[17] W. P. Dyke, Phys. Rev. **85**, 752 (1952).
[18] M. H. Nichols, following paper [Phys. Rev. **93**, 309 (1954)].
[19] Conyers Herring, *Structure and Properties of Solid Surfaces*, (University of Chicago Press, Chicago, 1953).

role in forming the structure on the 1980°K wire, although it must certainly be taken into account in a complete explanation of high-temperature structure. Herring[19] has pointed out that there is evidence that differential evaporation should tend to reduce rather than to enlarge the plateaus observed. He has also noted that at lower temperatures surface migration may be expected to become relatively more important, due to its lower activation energy.

Johnson[6] found no structure on single-crystal tungsten helices given ac heat treatment in vacuum sufficient to evaporate 17 percent of the wire diameter; from the known resolution of his observations, he concluded that there was no structure greater than a few thousand atom diameters in size on the crystals he studied. The structures observed here could just be resolved with optimum lighting. Although the plateaus are 10^{-3} cm wide, the structures are probably not much more than 5×10^{-5} cm in size and no larger than the upper bound set by Johnson. Thus it is possible that the plateaus on the helices were too small to be seen, particularly if Johnson did not use grazing incident illumination. However, Herring[19] suggests that the interchange of material between the surface of an *ideal* crystal facet and the region immediately below it may be so difficult as to inhibit the growth of such ideal facets. He points out that on Johnson's helices, facets would undoubtedly start as small elliptical spots of ideal orientation. Plateaus on long straight wires, on the other hand, are probably not exact (*hhk*) planes but vicinal planes, crossed by widely spaced transverse steps that would catalyze growth and dissolution. The shingle structure always found on long-wire plateaus is direct evidence of such transverse steps. Although Johnson found no plateau structure on vacuum-heated wires, he did report that flat surfaces appeared to develop in the (001) and (110) directions of crystals heated in lamp gas (86 percent argon plus 14 percent nitrogen) or in pure argon. He found that the (001) facets were larger than the (110) facets, whereas the reverse was observed in the present investigation. Evaporation is compensated for by condensation and produces no net material transport on a crystal in equilibrium with its vapor, as is approximately the case for a wire heated in a gas. On the other hand, the vacuum heated wire is subject to differential evaporation which could account for a (001) plateau of reduced width. Both the residual polishing marks and the maximum of diameter in the (001) direction can also be explained by differential evaporation. Using this hypothesis and the diameter plot of Fig. 3, Herring[20] has calculated an evaporation coefficient α_{001} of $\frac{1}{3}$ or $\frac{1}{2}$ or so. Using an analysis similar to that of reference 19, he has then shown that the (001) plateau width would be reduced to $\frac{2}{3}$ or $\frac{3}{4}$ of the value calculated[19] without differential evaporation, a result consistent with the observed plateau width.

The absence of a pronounced diameter maximum in the (110) direction must mean that there was almost no differential evaporation there. Herring has suggested two possible explanations.[20] First, the wire axis could be inclined to the (110) direction in such a way as to introduce steps close together on the (110) plateau but far apart on the (001). Second, the distance an atom can migrate before re-evaporation on the (001) plane may be substantially less than on the (110). In either case the evaporation coefficient α_{001} would be lower than α_{110}. This can be understood most easily by recalling that the principle of detailed balance requires the identity of the evaporation coefficient and the "sticking coefficient" for atoms impinging upon the surface from a vapor phase in equilibrium with the metal. An atom striking a plane surface from the vapor phase will be less likely to be incorporated into that surface if its chance to migrate to a step before being re-evaporated is small. This will occur if the steps are far apart or if distance migrated before re-evaporation is short; in either case α will be reduced.[21] Now from the last column of Table II it is seen that the wire axis lies very nearly in the 55° (112) plateau, since no shingling could be resolved on either the 55° or 235° plateaus. Then if the wire consists of one crystal and is truly cylindrical, the wire axis must make only a slightly larger angle with the 0° (110) plateau than with the 90° (001), so that steps on the (110) plateau should be spaced only slightly closer together than those on the (001) plateau. Close examination of the original photographs also shows (110) shingling only slightly more closely spaced than (001) shingling. Therefore, it seems that the second explanation is more plausible.

It is interesting to compare the rates of plateau growth with those obtained by Becker[16] for tungsten field emission points. Herring[22] has shown that under certain restrictions one can relate the times necessary to reach an equilibrium structure in two similar situations to the scale factor r of the two configurations. He has shown that if the mechanism is surface migration, the ratio of times to reach equilibrium will be r^4; if the mechanism is volume diffusion, the ratio will be r^3. Using these ratios and Becker's results, Herring[20] has calculated that it should take of the order of 5×10^5 days at 2400°K for sizeable plateaus to develop on 0.005-in. tungsten wires if the mechanism is surface diffusion. If the mechanism is volume diffusion, the time is reduced to about 3×10^3 days. Actually, plateaus are observed after only a few hours. Perhaps this discrepancy may also be due to inhibited plateau growth resulting from difficulty in interchange of material between the ideal crystal plane and the crystal bulk below that surface. As in the case of Johnson's

[20] Conyers Herring (private communication).

[21] See Appendix III of reference 10 for a discussion of smoothing by evaporation and related effects.

[22] Conyers Herring, J. Appl. Phys. **21**, 301 (1950).

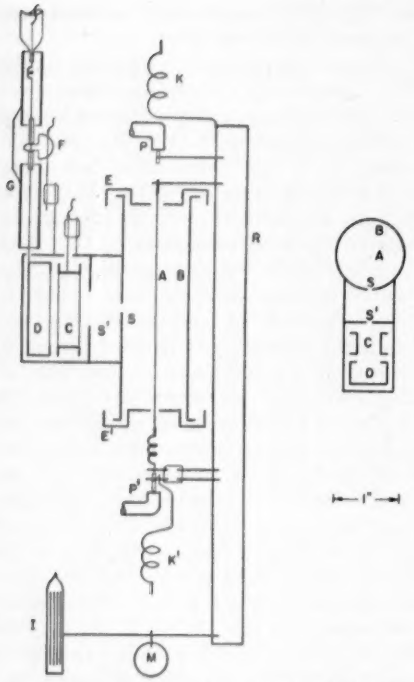

FIG. 4. Cross section and top view of tube used to measure the thermionic constants for different crystal directions. The crystal wire A is mounted in the rotor R so that it can be rotated in front of the slit S cut in the cylindrical anode B.

helices, Becker's field emission points must have facets starting in ideal orientation and hence would be subject to such inhibition, whereas the difficulty of interchange would not exist for the wires.

If the size of plateaus to be found on wires does vary with temperature in the range covered, the times used were not long enough to permit detection of such change. Of course, it is not likely that the wires ever reached an equilibrium structure, the rate of diffusion at low temperatures being too low, and evaporation disturbing any high-temperature equilibrium. The use of wires of smaller radius would be favorable for observations in shorter times.

IV. EQUIPMENT FOR THE THERMIONIC MEASUREMENTS

The tube used to make the quantitative thermionic measurements was originally designed and constructed by R. B. Nelson; it was first used to study crystallographic dependence of thermionic emission from tungsten single crystals by Nichols,[9] who rebuilt it to incorporate certain refinements. The tube was again rebuilt, with further modifications for the present measurements. Figure 4 is a schematic diagram of the metal parts of the tube as modified; the principal change from the tube as described in reference 9 is the addition of end guards E and E'. The slit S, 0.514×3.40 mm, is 3.1 times as small in area as that used by Nichols, giving an azimuthal resolution of 2.3° as compared with his resolution of 2.9°. The collimating slit S', 1.90×12.7 mm, does not serve to limit the beam.

The tube was pumped, outgassed, and sealed off by methods already described.[23] It is of interest that gas was released when the rotor R was first moved; care was taken to eliminate this gas before seal-off. A tantalum film getter along with a continuously operating Bayard-Alpert ion gauge[24,25] served to maintain a total pressure of about 1.5×10^{-9} mm of Hg for the entire period during which final measurements were made. "Emission decay" tests of 7 and 27 days duration at the outset and conclusion, respectively, of the final thermionic measurements showed no detectable change in emission in the (112) direction as measured at 1585°K. This indicates that the partial pressure of active gases, such as oxygen, was much lower than the total pressure, which probably consisted almost entirely of helium diffused through the glass.[26–28]

The Forsythe and Watson[29] temperature scale was used to determine the temperature of the middle section of the filament A. The 400-cycle alternating current used to heat the filament[30] was controlled by a thermionic regulator of the type first used by Ridenour and Lampson.[31] The rms value of the current was measured to better than 0.2 percent by comparison with a direct current producing the same heating effect in a 0.003-in. tungsten wire sealed in a tube containing hydrogen under a few mm of pressure. A Leed's and Northrup type-K potentiometer was used to measure thermal emf's and, with a standard 0.1-ohm resistor, the direct current. The best value of the wire diameter was obtained by use of an optical wedge interferometer; for the final thermionic data the average value from Fig. 3 is subject to a 0.2 percent correction due to evaporation[13] resulting from the heat treatment of Table I, G. The uncertainty in temperature due to 0.2 percent precision in diameter measurement is 3°K; that due to 0.2 percent precision in the current measurement is 2°K. Conduction cooling due to the finite length of the filament was observed in all Richardson plots for temperatures below about 1350°K, in good agreement with calculations based on tables given by Langmuir, MacLane, and Blodgett.[32]

Collector currents were measured by an Applied

[23] W. B. Nottingham, J. Appl. Phys. **8**, 762 (1937).
[24] R. T. Bayard and D. Alpert, Rev. Sci. Instr. **21**, 571 (1950).
[25] S. Wagener, Proc. Inst. Elec. Engrs. (London) **99**, Part III, 135 (1952).
[26] W. D. Urry, J. Am. Chem. Soc. **54**, 3887 (1932).
[27] W. H. Keesom, *Helium* (Elsevier, Amsterdam, 1942).
[28] D. Alpert, Phys. Rev. **89**, 902 (1953).
[29] W. E. Forsythe and E. M. Watson, J. Opt. Soc. Am. **24**, 114 (1934).
[30] In order to avoid the dc etch, as discussed in Sec. I, ac exclusively was used to heat the filament. Peak to peak temperature fluctuation at 800 cycles amounted to less than 1.5°K at 2000°K.
[31] L. N. Ridenour and C. W. Lampson, Rev. Sci. Instr. **8**, 162 (1937).
[32] Langmuir, MacLane, Blodgett, Phys. Rev. **35**, 478 (1930).

Physics Corporation vibrating reed electrometer[33] with one precision wire-wound and six Victoreen resistors. The anode voltage was supplied by a well-filtered full-wave rectifier circuit, *VR*-tube regulation being used for Richardson plots. Anode voltage was measured with the type-*K* potentiometer. All other electrode potentials were battery supplied.

Preliminary to actual use of the tube and before the end guards *EE'* were added, a complete set of operating characteristics was obtained to evaluate the effectiveness of, and correct operating point for, each of the auxiliary electrodes. Two forms of anomalous behavior, not previously reported by Nichols in his earlier work with the same tube, were discovered by their influence on the form of the operating characteristics. These spurious effects were more noticeable in the present investigation because (a) the slit *S* was reduced in cross section by a factor of three, and (b) operating characteristics were taken with the crystal oriented to measure a minimum of thermionic current. The orientation used by Nichols is not known, but judging from the nature of his characteristics and the order of magnitude of his collected currents,[9,34] it was probably not a direction of minimum emission.

Evidence of the first anomalous effect was found in the operating characteristic of the guard ring G. The proper guard ring potential required to shield the collector lead from stray electrons escaping from the open end of anode *B* (without end guards) was determined by plotting the collector current I_c *vs* the guard ring potential E_g with respect to the anode *B*. The collector was held at anode potential. Plots for crystal directions other than the minimum (110) were well-behaved and similar to that obtained by Nichols.[34] The plot taken with emission from the (110) direction (Fig. 5a) has two distinguishing features in addition to the expected peak in the negative vicinity of zero E_g. First, the current collected in the most negative guard ring region, although independent of E_g, has a much greater dependence on anode-cathode voltage E_p than does the current collected in the positive guard ring region. Secondly, there is an almost linear dependence of I_c on E_g for positive E_g. The first feature appears to be due to photoelectric emission from the guard ring, produced by x-rays of 1 to 3 kv from anode *B*, and collected by the collector lead due to the favorable field between it and the guard ring. The reverse phenomenon, photoelectric emission from the collector lead when E_g is positive, is not so pronounced due to the much smaller surface area of the collector lead. One cannot account satisfactorily for this first feature of Fig. 5a by mechanism (photoelectric effect, stray electron currents, field emission, ion currents) other than x-ray photoelectric emission. The second feature of Fig. 5a, the linear decrease of I_c with increasing E_g, appears to be due to electron emission from the glass surface of the insulator that supports the collector, particularly from the Pyrex wetting glass which provides a thin covering for the collector lead over several millimeters of its length. This emission is most likely secondary emission arising from bombardment of the glass by stray primary electrons, although there may also be some photoelectric emission produced by x-rays. Since the nearest electrode to this glass is the guard ring itself, the emission serves to charge the surface of the glass to the potential of the guard ring. From the slope of the curves of Fig. 5a, the resistance of the leak is 5×10^{13} ohms, a value consistent with the resistivity of Pyrex, 10^{14} ohm cm, and the dimensions of the wetting glass. The anode potential dependence of I_c observed for a constant positive E_g is very nearly in agreement with the Schottky mirror image theory, as might be expected if the Schottky dependent emission has a small constant superposed leakage current. Further evidence of dielectric charging is given by long times required for reaching equilibrium after changing conditions in the measuring tube, especially if a change in E_g is involved. Nichols reported a similar time effect for measurements in the (110) direction.[34] A volt-ampere characteristic for the guard ring also was consistent with the above explanations of both features of the first anomalous effect.

The choice of a completely satisfactory operating point on the curves of Fig. 5a is of course impossible. However, since the slope on the E_g positive side is significant only for low temperatures *and* in the (110)

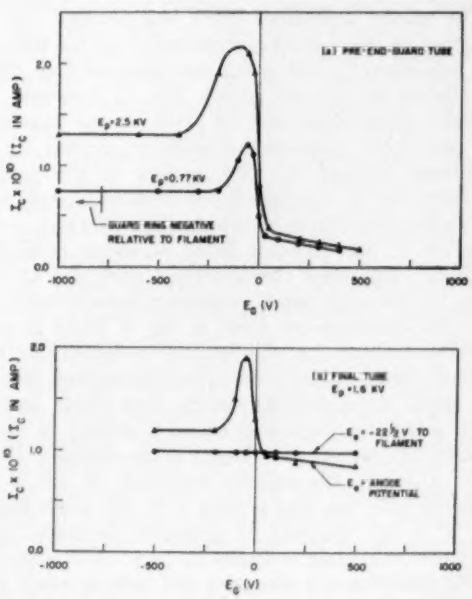

FIG. 5. Guard-ring characteristics before and after elimination of first anomalous effect. Data taken in (110) direction at 1690°K.

[33] Palevsky, Swank, and Grenchik, Rev. Sci. Instr. **18**, 298 (1947).

[34] M. H. Nichols, Massachusetts Institute of Technology, Ph.D. thesis, 1939 (unpublished).

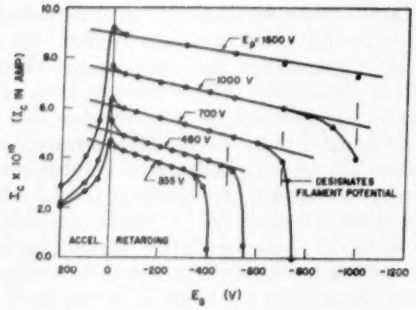

FIG. 6. Suppressor characteristics for final tube showing second anomalous effect. Data taken in (110) direction at 1805°K.

direction, and since the x-ray photoelectric current is large only for negative E_g, a value of 180 v positive was chosen for the unmodified tube as used to obtain the "pre-end-guard" data. Before taking the final data, the end guards EE' were installed as shown on Fig. 4. Their presence and operation at 22.5 v negative with respect to the filament served to eliminate both radiation and stray electrons from the vicinity of the guard ring, as can be seen from the guard-ring characteristic of Fig. 5b. Their operation at anode potential, however, does not eliminate the glass charging effect, indicating that it is at least partly due to stray electrons from the end of the anode. No distortion was introduced in the field at the slit by the presence of the end guards, as evidenced by constant collector current for a maximum emission direction when the end guards were varied separately or simultaneously from anode to filament potential.

The second anomalous effect was revealed by its influence on the operating characteristic of the secondary suppressor C. This effect was observed even after installation of the end guards. For all directions of emission except those near the (110) minima there was no observable variation of collector current I_c *vs* suppressor voltage E_s with respect to the anode, provided that E_s was sufficiently negative to repel and return slow secondaries generated at the slits and at the collector D, the collector being at anode potential. Nichols'[34] suppressor characteristics were also of this form. On the other hand, suppressor characteristics for the (110) direction are given in Fig. 6. There is considerable evidence that this behavior is due to collection through the slit system of secondary electrons emitted at the anode B, secondaries coming from anode regions upon which the impinging emission density is much greater than the emission density in the direction of the slits. Although the angular resolution of slit S to emission from the wire is only 2.3°, the much larger slit S' is not designed to collimate the beam, and electrons proceeding at angles up to 6° or 7° with the radial direction can penetrate the slits to reach the collector D. Moreover, the vertical resolution is very poor indeed, due to the length of S and emphasized by the excessive length of S'. A typical electron path is shown in Fig. 7. The trajectory was calculated by a graphical stream plotting method for an electron leaving the anode at 5° from the normal, with one-half of the primary electron energy. This electron originated at 120° in azimuth, measured from the slit. The curves of Fig. 6 are essentially integrated energy distribution curves of the current through the slits. The form of the curves indicates an energy distribution similar to that known for secondary emission.[35–37] There is a low-energy peak, an almost constant distribution over medium energies, and a large full-energy peak. Indeed, if it were not for the full-energy component of the secondary emission it would be a relatively simple matter to use the suppressor characteristic to subtract out all but the full-energy thermionic emission. The curves of Fig. 6 indicate that the spurious effect is at least as pronounced at low anode voltages as at higher ones. Although not too much is known about primary energy dependence of secondary emission components, the total secondary emission ratio is known to have a maximum at about 600 v for tantalum.[38] Suppressor curves for various temperatures show the effect to be relatively larger at lower temperatures. This would be expected since the surfaces producing the abundant primary emission responsible for secondary emission are of lower work function than the (110) surface. Elimination of the second anomalous effect was not undertaken. The technique of correction used to obtain (110) data will be described in Sec. V. For directions other than the (110), the suppressor electrode was operated at −540 v with respect to the anode, except for retarding potential plots.

As long as the collector D was positive with respect both to the filament and to the suppressor, there was no dependence of collector current upon collector voltage. Except for retarding potential plots the collector was operated at anode potential.

V. THERMIONIC DATA FROM SINGLE CRYSTAL OF TUNGSTEN

Two complete sets of thermionic data were obtained: (a) the pre-end-guard data, subject to both anomalous

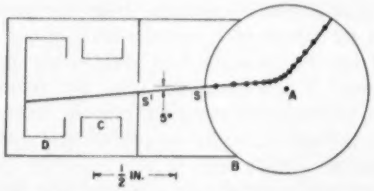

FIG. 7. Graphically computed path of a secondary electron emitted from anode so as to enter slits SS' to be collected at D. The electron is emitted with one-half primary energy at 5° to the radial direction.

[34] K. G. McKay, *Advances in Electronics* (Academic Press, Inc., New York, 1948), Vol. 1.
[35] Erick Rudberg, Phys. Rev. **50**, 138 (1936).
[37] L. J. Haworth, Phys. Rev. **48**, 88 (1935).
[38] L. R. Koller, Gen. Elec. Rev. **51**, 33, 50 (1948).

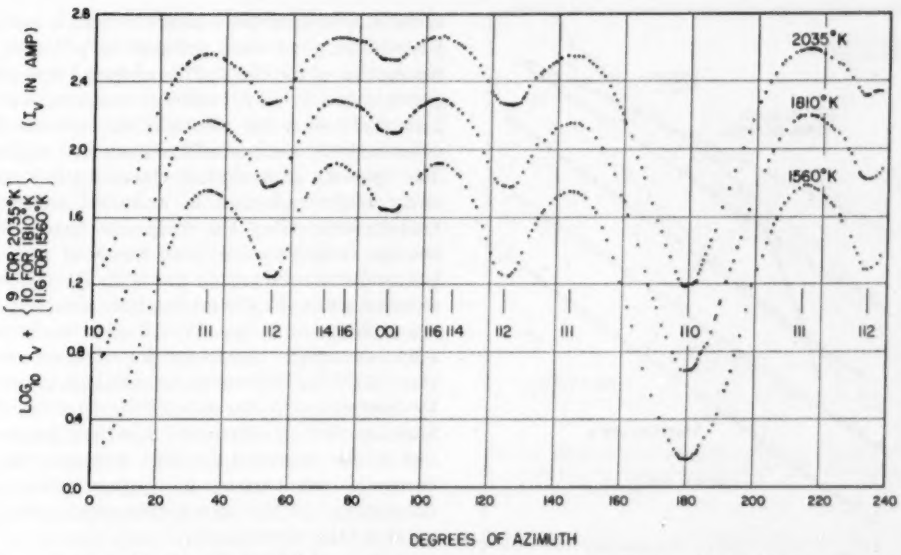

FIG. 8. Emission *vs* azimuth; $E_p = 1600$ v, corresponding to surface gradient of 5.5×10^4 v/cm.

effects, and (b) the final data, with end guards installed to eliminate the first anomalous effect. The pre-end-guard data were taken with the tube basically as used by Nichols; the results show good agreement with those he obtained. The first anomalous effect is important only in directions of low emission, as will be seen from Richardson plot comparisons. The second anomalous effect, not eliminated in either set, seriously affects the measurements only in the (110) direction, where the data have been partially corrected, as will be described. At the conclusion of the final run the wire was flashed for one min at 3030°K to check reproducibility of the data. For no crystal direction did emission after the flash differ by more than 5 percent from the pre-flash value.

In Fig. 8, polar plots of emission *vs* azimuth about the wire were obtained by rotating the crystal wire before the slit system. Assignment of crystal directions is in agreement with previous work.[4,9] The distinguishing feature of the polar plots is the appearance of definite "flats" in the directions of minimum emission. This corroborates the visual observation of plateaus on the wire, as discussed in Sec. II. Nichols' polar plots[9] did not show these flats. This can be due to: (1) an insufficient number of points in his curves together with slightly poorer resolution, or (2) a different heat treatment of his crystal. Although the heat treatment Nichols gave his crystal after repolishing and before taking thermionic data was sufficient to give reproducible results, it was quite mild compared to that given here. However, since plateaus have been optically observed on wires given only very mild heat treatment (see Sec. II), the more probable explanation is (1). The data of Fig. 8 were taken with the final tube; the emission flats were found on preliminary runs, but there had been heat treatment sufficient to produce 2.5 percent wire diameter reduction due to evaporation before the first sufficiently detailed run was made. It should be noted that currents smaller than that in the (112) direction are not reliable, owing to the second anomalous effect. For this reason, it is also not certain that the (110) plateau width derived from Fig. 8 is accurate.

Schottky plots of emission *vs* anode-filament potential E_p were taken at two temperatures for each direction corresponding to a maximum or minimum of emission. Illustrative final tube curves are shown in Fig. 9. The pre-end-guard data did not differ significantly from the final data except in the (110) direction, where the former were difficult to obtain due to glass charging effects. The slope S of the Schottky line (the coefficient of $E_p^{\frac{1}{2}}/T$) was used to reduce the high field emission to zero field conditions. Agreement of slopes for the two temperatures was within 5 percent. Average values are tabulated for each direction in Table III. The numerical values of slope are considered to be within experimental error of the mirror image theoretical value for a uniform surface, except in the (110) direction. The (110) data will be discussed separately below. Evidence is seen on the Schottky lines of the periodic Schottky deviation.[39–42] The upper voltage limit of 2800 v, imposed by the onset of field emission from tube parts, limited the voltage range of the tube to about one cycle of the deviation, making it difficult to choose a reference line for the fluctuation.

[39] R. L. E. Seifert and T. E. Phipps, Phys. Rev. **56**, 652 (1939).
[40] D. Turnbull and T. E. Phipps, Phys. Rev. **56**, 663 (1939).
[41] W. B. Nottingham, Phys. Rev. **57**, 935 (1940).
[42] Munick, LaBerge, and Coomes, Phys. Rev. **80**, 887 (1950).

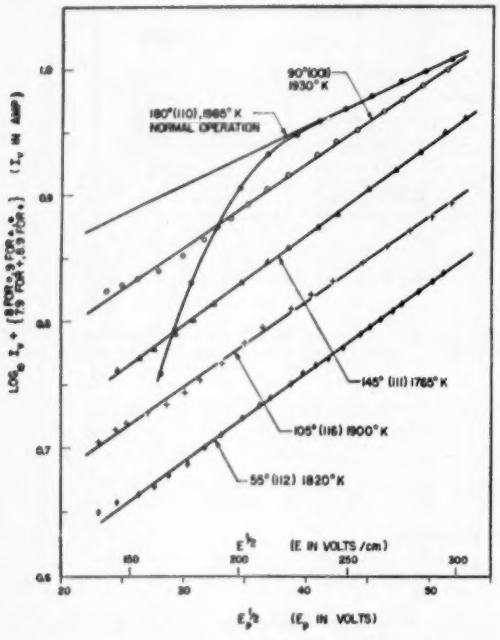

FIG. 9. Illustrative Schottky plots.

No attempt was made to study the periodic deviation due to the limitations of the tube. Directions other than the (110) show no deviation from the Schottky line, aside from the periodic deviation, for voltages higher than 750, corresponding to a gradient of 26 000 v/cm. A simple calculation shows that the upward deviation in the (112) and (001) directions at this point is clearly the result of failure of projection properties of the tube due to thermal velocities. Similar deviation occurs only at even lower potentials for the high-emission directions, as might be expected in view of the broad nature of the peaks of emission (Fig. 8). Patchiness, either small or large scale, can produce only downward deviation in the (112) and (001) directions.

Richardson plots of emission *vs* temperature were taken in each minimum and maximum emission direction. The apparent emission constants for uniform and for patchy surfaces are defined and discussed by Herring and Nichols.[10] The experimental constants will be doubly starred because there is insufficient evidence to show that the exposed surfaces are not patchy. Illustrative Richardson plots for the final and pre-end-guard tubes are given in Fig. 10. Complete results from both tubes appear in Table III, along with the values obtained by Nichols.[10] Difference between the final and the pre-end-guard data is a measure of the first anomalous effect. Except in the (110) direction the effect is seen to be small.[43] Data from high-emission surfaces were essentially unaffected while low-emission Richardson plots were reduced in φ^{**} and A^{**} by elimination of the effect. Pre-end-guard operation, with guard ring at +180 v, is subject to the reduced anomalous effect of x-ray photoelectric emission from the collector lead, which should produce just such an effect. The spurious photoelectric emission produces a relatively larger reduction in measured current at low temperatures, since the x-ray are generated by an average (relatively low work function) emission. The low-temperature droop in the (110) Richardson line is a consequence of collector insulator charging and leakage as discussed in Sec. IV. All other Richardson lines show downward departure for temperatures below about 1350°K. This is due to reduction in temperature by conduction to the ends of the wire (see Sec. IV). Slight systematic downward departure from a straight line is also observed for high temperatures. Careful check of the current measuring apparatus leaves uncertainty in the temperature scale as a possible source of this deviation.[44]

Data for the (110) direction, *taken in the same way as for other surfaces*, are in good agreement with those obtained by Nichols.[9] Both the final tube and the pre-end-guard tube gave Schottky plots with slightly lower than average slope and with definite departure from the Schottky line at about the same surface gradient as observed by Nichols. The final tube Schottky plot for the (110) direction is given along with valid plots for other directions in Fig. 9. The pre-end-guard thermionic constants obtained by normal operation of the tube (Tables III and IV) are in good agreement with Nichols' constants. It will be recalled that the crystal used was selected with the purpose of eliminating the shingle structure on the (110) surface. Although shingling was not in fact completely missing, the degree of shingling was greatly reduced from that encountered by Nichols. If the shingling were indicative

TABLE III. Thermionic constants for the different crystal directions. Both final and pre-end-guard values are tabulated. Nichols' results, reference 10, are given for comparison. The precision of φ^{**} is about 0.5 percent, that of A^{**} about 10 percent; S is the Schottky slope.

Direction	Final φ^{**}	Final A^{**}	Final S	Pre-end-guard φ^{**}	Pre-end-guard A^{**}	Pre-end-guard S	Nichols φ^{**}	Nichols A^{**}
35° (111)	4.39	54	12.8	4.38	55	11.9		
145° (111)	4.37	47	12.8	4.36	49	12.0	4.39	35
215° (111)	4.38	54	12.3	4.39	65	10.6		
55° (112)	4.65	118	12.8	4.68	163	12.9		
125° (112)	4.64	110	13.2	4.68	150	12.5	4.69	125
235° (112)	4.65	133	11.5	4.66[a]	153[a]	10.8		
75° (116)	4.29	40	12.8	4.31	48	12.2	4.39	53
105° (116)	4.29	40	12.3	4.30	46	12.0		
90° (001)	4.52	105	12.7	4.52	119	12.4	4.56	117
180° (110)[b]	4.58[b]	8.0[b]	9.1[b]	4.66[b]	11.8[b]	10.4[b]	4.68[b]	15[b]
Theoretical			11.2			11.1		

[a] Unreliable data due to deterioration of vacuum over long interval between remainder of pre-end-guard tube data and this measurement.
[b] Anomalous behavior invalidates (110) data obtained under normal operating conditions. True (110) current smaller than these figures indicate.

[43] The location of the pre-end-guard Richardson lines *above* the final lines is probably due to error in pre-end-guard wire diameter, obtained by 0.61 percent correction for evaporation from final line value; see reference 14.

[44] Temperature dependence of work function can produce curvature of the Richardson line. It will be recalled, however, that a *linear* temperature dependence of φ for a *uniform* surface will produce no curvature. Although patchiness can produce curvature, it will always be concave upward.

of patchiness on the (110) surface and if that patchiness were responsible for departure from the Schottky line, one would expect higher φ^{**} and A^{**} from the present measurements, rather than the slightly lower pre-end-guard values obtained. When the first anomalous effect was eliminated in the final tube, the constants were affected in the same way as were those of other low-emission surfaces, but to a larger degree. The more important second anomalous effect, due to secondary emission, can be partially eliminated by using the suppressor C as a retarding barrier. By extrapolating suppressor characteristic curves (Fig. 6) to the filament potential one can eliminate all but full-energy electrons to the collector. The "extrapolated" Richardson plot obtained in this way (Fig. 11) has emission constants $\varphi^{**}=4.72$, $A^{**}=9.7$, which set a definite upper limit for the true (110) emission. An estimate of the value of the true (110) work function can be made under two assumptions: (a) that all of the secondary emission current has the same effective work function, and (b) that, as has already been found for the similar (112) and (001) surfaces, $A^{**}_{(110)}\approx 120$. Under assumption (a), the difference between the left and right extrapolations of the suppressor characteristic is taken as a measure of the secondary emission current, and an effective work function, 4.42 v, is obtained from a "Richardson plot" of this current (Fig. 11). Then by a method of successive approximations, just enough spurious current with this effective work function is subtracted to leave a Richardson plot with $A^{**}=120$. The apparent work function of the estimated (110) line so obtained is 5.26 v. The agreement between the spurious current effective work function, 4.42 v, and the average work function of doped polycrystalline tungsten,[3] 4.45 v, suggests that the secondary emission as obtained here is indeed dependent upon a high-field average over all of the crystal, supporting assumption (a). To illustrate the magnitude of concave upward curvature expected in the "extrapolated" Richardson plot,[10] straight 5.26 and 4.42 v Richardson lines were added to give a perfect match to the experimental curve at $10^4/T$ of 5.3 and 6.2. Points on this computed line are plotted as triangles in Fig. 11; it is seen that the curvature produces a negligible deviation from the straight line in the range of temperature for which data are available. Due to dependence of the secondary

TABLE IV. Summary of (110) data.

Description	φ^{**}	A^{**}
Nichols' values, reference 10	4.68	15
Pre-end-guard, normal operation	4.66	12
Final tube, normal operation	4.58	8.0
Final tube, extrapolated	4.72	9.7
Pure spurious current, effective slope	4.42	
Estimate by subtraction of spurious current to give $A^{**}=120$	5.26	120
Contact potential (implicitly assuming $A^{**}=118$)	5.21	118

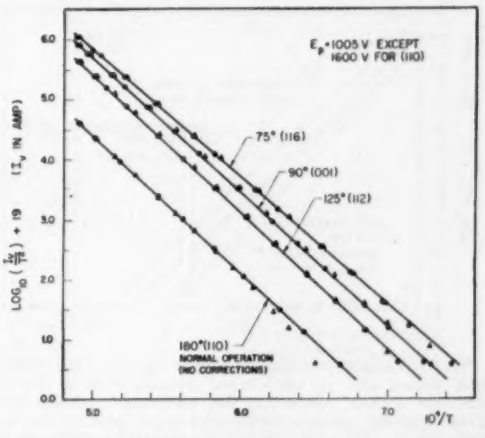
E_p = 1005 V EXCEPT 1600 V FOR (110)
75° (116)
90° (001)
125° (112)
180° (110) NORMAL OPERATION (NO CORRECTIONS)
$\log_{10}(\frac{i_s}{T^2}) + 19$ (i_s IN AMP)
$10^4/T$

FIG. 10. Illustrative Richardson plots of pre-end-guard data ▲ and corresponding final tube data ●. The Richardson lines are drawn for the final data.

emission process on primary energy, no data significant in the usual sense could be obtained from a (110) Schottky plot by extrapolation procedures. The plot shared the rapid downward trend at low voltages which marked the original unextrapolated Schottky curves, but for higher voltages it became almost constant; 745 v was chosen as the optimum anode-filament potential difference consistent with satisfactory projection properties of the tube. Since no meaningful Schottky slope was available, the theoretical slope of 11.2 was used to correct the data to zero field values. A summary of various apparent thermionic constants for the (110) direction appears in Table IV.

Although the tube as described was not designed for measurements using retarding fields, operation with the collector within a few volts of the filament potential permitted the taking of retarding potential curves for the various crystal directions. The anode was held at 1005 v to maintain the projection properties of the tube. The maximum deviation from the mean filament potential due to ac drop amounted to 0.13 v. Although there was no completely satisfactory operating potential for the suppressor electrode C, the results to be reported were taken with the suppressor 3 v positive with respect to the filament, the value which gave maximum collected current. Data taken with the suppressor at collector potential or at 1.5 v negative with respect to the collector were essentially the same. Retarding potential curves for one each of the (111), (112), (116), (001) and (110) surfaces are given in Fig. 12. As expected, all curves, except the (110), merge for strongly retarding fields.[45] The slope for strongly retarding fields represents a temperature of 2090°K,

[45] The fact that the curves, other than the (110), do not coincide precisely for strongly retarding fields is probably due to slight differences in temperature among the runs. However, the short time stability of the filament current regulator did not permit significant temperature drift during one retarding run.

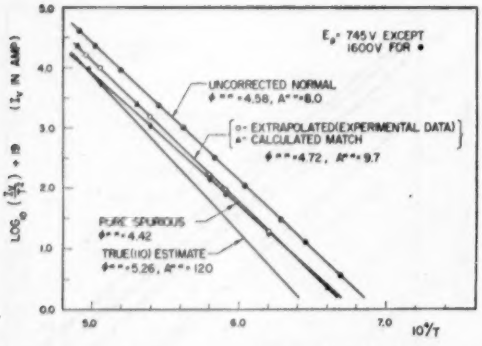

FIG. 11. Final tube Richardson plots for the (110) direction. Data plotted are: (1) uncorrected normal, taken as for other directions without correction for secondary emission effect; (2) extrapolated, giving a conservative upper limit for the "true" (110) current; (3) pure spurious, representative of the spurious secondary current alone; (4) estimate of the true (110) current.

whereas the value computed from the filament current and the Forsythe-Watson temperature scale is 1990°K. The saturation values have been designated somewhat arbitrarily, since the values indicated represent only about 55 percent of the current collected under normal accelerating conditions at the same temperature. However, the fraction collected is the same for each direction (except for the (110), where the fraction is 34 percent of the extrapolated Richardson current). The remainder of the slit current is divided between the suppressor C and the anode in the unfavorable electron optical configuration. That the (110) saturation value is low indicates more complete elimination of spurious secondary current under retarding conditions. Except for the (110) direction, no new information is to be had from contact potentials from pairs of retarding potential curves,[46] since the saturation current is equivalent to a Richardson line point, and the curves have the same tangent for strongly retarding fields. But in view of the reduced (110) saturation value, it is interesting to obtain a contact potential for that direction. The value obtained with the 55° (112) curve, taking the difference in abscissas of points A and B of Fig. 12, is 0.64 v, giving $\varphi_{(110)}{}^{**}=5.29$ v. Point B is taken on the common tangent since the (110) curve must merge with that curve for sufficiently retarded conditions. This φ^{**} value[47] contains the implicit assumption that $A_{(110)}{}^{**}=A_{(112)}{}^{**}$. Since $A_{(112)}{}^{**}=118$, this is equivalent to assumption (a) of the previous paragraph. However, the comparison does not involve a subtraction of spurious current; thus it gives a more reliable upper limit for the true (110) emission at this temperature. The shape of the retarding potential curves can sometimes give information about the patch nature of the emitter.[48] Patchiness produces a breaking away from the corner of a retarding potential curve. Unfortunately, the departure from an ideal plane parallel configuration also causes a rounding off of the corner, making the observation of patch effects more difficult. From the shape of the curves of Fig. 12, it is clear that the ac filament drop does not make a large contribution to the rounding. Although the curves are all of very nearly the same shape, that for 55° (112) shows the least rounding at the breakpoint. In Fig. 13 the 75° (116) curve is compared with the 55° (112) curve by matching saturation currents and currents for most retarded conditions. Excluding the (110) direction which will be discussed in the next section, the comparison is typical, the poorest match occurring between the 235° and 55° (112) directions. To give some idea as to what sort of rounding off patchiness might produce, two-patch curves were computed to match the 75° (116) emission. The 55° (112) curve was taken as characteristic of a uniform surface and proper proportions of current from two such curves, displaced in abscissas, were added to obtain a retarding potential curve for the two-patch surface. Using Fig. 12 of reference 10, patches with $A=120$ and no work function temperature dependence were chosen to give the

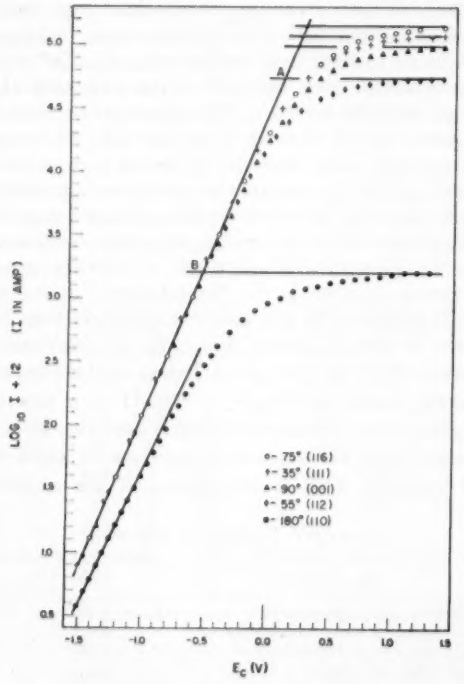

FIG. 12. Retarding potential curves for representative directions (final tube data); $E_p=1005$ v.

[46] As pointed out by Herring in private communication. The view taken in the U. S. Office of Naval Research Report N6-onr-24433 (unpublished) on this matter is erroneous.

[47] One notes that the same values are obtained by arbitrarily taking $A_{(110)}{}^{**}=118$ and choosing $\varphi_{(110)}{}^{**}$ so that Richardson's equation will give the saturation (110) current of Fig. 12.

[48] See reference 10 for a discussion of retarding potential plots for patchy emitters.

experimentally observed emission constants $\varphi^{**}=4.30$, $A^{**}=40$. From reference 10 it is seen that a $\delta\varphi \gtrsim 0.4$ v is required to give $A^{**}=40$. For $\delta\varphi=0.4$, 10 percent of the surface must have $\varphi_1=4.20$ and the remainder $\varphi_2=4.61$. For $\delta\varphi=1.0$, one may have 33 percent $\varphi_1=4.30$ and the remainder $\varphi_2=5.30$.[49] Retarding curves for these two composite surfaces are plotted along with the experimental curves in Fig. 13. The two-patch curves do show a larger falling-off in the vicinity of the breakpoint than does the experimental curve. However, it is not certain that a larger number of patches might not give smaller departure in form from the uniform surface retarding potential curve.[50]

VI. DISCUSSION OF EMISSION MEASUREMENTS

Since the first anomalous effect produced only small changes in the thermionic constants, and since the second effect was negligible (except in the (110) direction), it is reasonable to compare data (Table III) with those obtained by Nichols.[9] In general, the precision limits overlap. However, $\varphi_{(116)}{}^{**}$ is 0.09 v lower than Nichols' value. It is possible that the difference is due to different wire heat treatments before measurements were made. Although the treatment Nichols gave his crystal after repolishing and before taking thermionic data was sufficient to give reproducible results, it was quite mild compared to that given here. It is likely that the treatment of Table I, CDEF produced material transport responsible for the difference in $\varphi_{(116)}{}^{**}$. (111) constants are in good agreement, showing that the effect, if any, of the dc etch on Nichols' crystal apparently did not extend to the (111) direction.

It has been shown that (110) data taken in the same way as for other surfaces are seriously affected by the second anomalous effect even after elimination of the first anomalous effect. Indeed, both the break in the (110) Schottky plot and the very low value of $A_{(110)}{}^{**}$ reported by Nichols and observed here are probably entirely due to this secondary emission effect, rather than to patchiness on the (110) surface. The shape of the (110) retarding potential curve of Fig. 12 also is undoubtedly influenced by remaining spurious current. However, when the retarding barrier at the collector becomes higher than the emitter surface barrier of any patch, one would expect the true (110) current to merge into the common line, as was the case for all other surfaces. The fact that it lies *below* that line for the most retarded potentials used clearly indicates a deficiency rather than a surplus of current. This can be explained by (1) the presence of a high work function patch on the (110) surface, (2) a large reflection

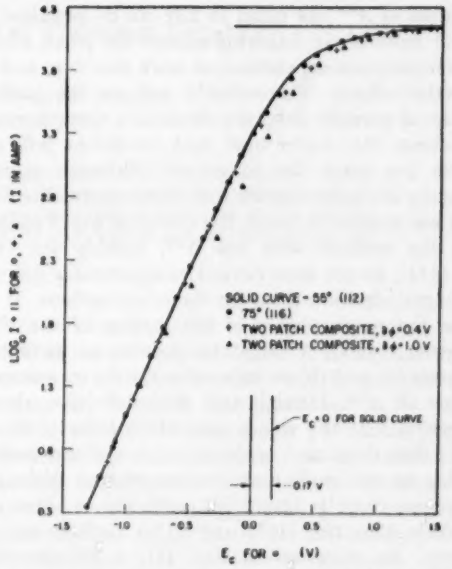

FIG. 13. Comparison of 75° (116) and 55° (112) retarding potential curve shapes. Calculated two-patch curves matching the 75° (116) emission are also plotted.

coefficient for the (110) surface, or possibly (3) a failure of projection properties of the tube in the (110) direction. The (110) curve has not yet joined the common tangent at -1.5 v in Fig. 12. The (112) curve joins the tangent at about -0.4 v. If (1) is the explanation, this would indicate the existence of a patch on the (110) surface more than 1.1 v higher than $\varphi_{(112)}{}^{**}$, at least 5.7 v and likely as high as 6.0 v. Since a Schottky plot could not be obtained for the (110) direction, (2) and (3) cannot be ruled out. However, tangential patch fields near the (110) direction would have to be much higher than those near the (112) direction for failure of projection properties since (112) Schottky plots showed satisfactory projection properties until thermal velocities became important. Assuming satisfactory projection in the (110) direction, one can say from the saturation value of the (110) retarding potential plot that the (110) emission at 1990°K is no larger than would be given by $\varphi^{**}=5.29$, $A^{**}=118$. There is no justification for that arbitrary choice of A^{**} except the similarity among the (110), (112), and (001) surfaces. All of them have plateaus with shingling, and each of the latter two has A^{**} near 120. From the extrapolated Richardson plot one can say that the (110) emission at any temperature within the range covered is no larger than that given by $\varphi^{**}=4.72$, $A^{**}=9.7$, although this certainly includes some of the spurious current.‡

[49] Although Fig. 12 of reference 10 shows two ways of obtaining the desired φ^{**} and A^{**} with $\delta\varphi=1.0$ v, the one with the higher percentage of low work function patch has been chosen. This avoids excessive curvature of the Richardson plot for the two-patch surface.

[50] Two-patch curves were also computed to match the 35° (111) emission. The results were very similar to those for the 75° (116).

‡ *Note added in proof.*—Drechsler and Müller, Z. Physik **134**, 208 (1953), M. Wilkinson, J. Appl. Phys. **24**, 1203 (1953), and Dyke, Trolan, Dolan, and Grundhauser, J. Appl. Phys. **25**, 106 (1954), have recently reported field emission $\varphi_{(110)}$ values of 5.70, about 4.7, and about 5.2 v, respectively. Drechsler and Müller's value is derived from measurements first reported in reference 15;

Values of A^{**} not equal to 120 can be explained by one or more of the following effects: (a) patch effects, (b) temperature dependence of work function, and (c) reflection effects. Theoretically and on the basis of studies of periodic Schottky deviations there is reason to believe that under high field conditions reflection effects are small for tungsten.[10] Although periodic Schottky deviations for the individual crystal directions were not studied in detail, the curves of Fig. 9 indicate that the surfaces with low A^{**}, notably the (116) and (111), do not show deviations appreciably different in magnitude from those for the other surfaces. If one knew the patch nature of the surface of the single tungsten crystal, it might be possible to distinguish between (a) and (b) as responsible for the experimental values of A^{**}. Herring and Nichols[10] have already suggested that the values near 120 for the (112) and (001) directions are consistent with uniform surfaces having no net work function temperature derivative. However, there is insufficient evidence to show conclusively that the (111) and (116) surfaces are not patchy. As discussed in Sec. III, a hill-and-valley surface is a possible quasi-equilibrium structure on a relatively large crystal. There is, on the other hand, no evidence other than the variation in A^{**} that the (111) and (116) surfaces *are* patchy. Microscopic examination of the crystal surface, capable of resolution to about 1μ, showed no structure except the plateau and shingle structure already discussed. The plateaus appear to be associated with simple facets exposed on the single crystal; the fact that the range in degree of shingling encountered had no effect on the emission would indicate that such structure does not represent a significant patchiness (at least in the (112) and (001) directions). Tangential fields due to patches of 1μ linear dimension and 0.5 v work function difference would be of the order of 0.5×10^4 v/cm and should easily affect the slope of a Schottky line at 2.0×10^4 v/cm, the lower limit of the Schottky plots of Fig. 9. The experimental Schottky lines were straight and had essentially theoretical slope for fields up to the limiting value of 1.0×10^5 v/cm. If the region plotted lies entirely below the anomalous Schottky region due to a fine scale patchiness (case 2 of reference 10), the patch fields would have to be of the order of 10^6 v/cm. Again assuming a 0.5 v work function difference, this would require a patch dimension of the order of 50A, only one order larger than the lattice spacing. Finally, the retarding potential plots (excluding the (110) plot) showed no significant rounding of the corner due to a patchy emitter. However, the calculated two-patch curves show that for the unfavorable electrode configuration used, the shape of the retarding potential curve is not a sensitive test for patchiness. Thus it appears that visual observation and straight Schottky lines of theoretical slope rule out large and medium scale patchiness, leaving a small scale effect possible.

Especial gratitude is due to M. H. Nichols, who suggested the problem, for his continued interest and inspiration, for stimulating discussions, and for his many suggestions. I am indebted to Conyers Herring for much of the interpretation of the surface structure observations and for reading the manuscript. I am also grateful to the Standard Oil Company of California for the Standard Oil Fellowship, during the tenure of which part of this work was done.

they measured the (110) current through a small anode aperture. The other two values were obtained by phosphor brightness measurements. Dyke *et al.* took special precautions to minimize light scattering into the faint (110) areas.

Growth and Surface Properties of Tantalum Crystals*

M. H. NICHOLS†
California Institute of Technology, Pasadena, California
(Received November 9, 1953)

A method consisting of a moving temperature gradient for producing centimeter-long crystals in commercial tantalum wire is described. Based on one run, the method was less successful on commercial molybdenum wire. Thermionic projection tube patterns from the resulting crystals were qualitatively the same as that from tungsten. Plateau-like structure similar to that of tungsten was observed on both the tantalum and molybdenum crystals.

THE three most refractory metals are tungsten, tantalum, and molybdenum. Since these three all form body-centered cubic crystals, they might be expected to display similar crystallization, surface-structure, and electron-emission properties. Of these three metals, only the properties of tungsten have been investigated extensively.

I. CRYSTAL GROWTH TECHNIQUE FOR TANTALUM

In order to study the thermionic properties of crystals, it is most convenient to use crystals the order of a centimeter in length in fine wires.[1] In the case of tungsten wires, which are widely used for incandescent lamp filaments, additives have been developed which promote large grain growth[2] so that only a simple heating schedule of the raw wire is necessary to produce crystals centimeters long occupying the entire wire cross section.[3] In the cases of tantalum and molybdenum, there has been no development similar to that of tungsten and only "commercially pure" wire is available. The tantalum wire used was obtained from the Fansteel Metallurgical Corporation and was designated T71-Q25014A-No. R8130-8 dated March 13, 1951. It is well known that the crystal growth properties depend on the purity of the metal and on the mechanical strain in the wire.[2,4] The writer has been informed that the Fansteel wire "conforms with our specifications for purity which designate limits; namely, iron 0.03 percent max, carbon 0.03 percent max, and tantalum 99.9 percent min." "Variations between lots may be expected to be small" and "no detailed analysis of the specific lot is available."[5] Mrowca[4] has reported that in the case of tantalum ribbon, a cold stretch resulting in about 2 percent elongation produced the largest crystals.

All the wires were 0.005 inch in diameter and were polished to remove the die marks.[6] With the exception of one run on tantalum in which helium was used as a filler gas, all runs were made in tubes continuously pumped by mercury vapor pumps with no stopcocks on the high-vacuum side. After baking at 500°C, a vapor trap between the tube and the pumps was refrigerated by liquid nitrogen during the runs. The Langmuir-Malter temperature scale was used.[7] All heating was done on ac to avoid dc etching.[8]

As a preliminary test, a tantalum wire 30 cm long was first given a cold stretch of 2 percent and then mounted in an electron projection tube as indicated in Fig. 1(a).[9] A heating schedule similar to that described by Robinson[3] was used consisting of 21 hours at 2100°K, 26 hours at 2300°, 7 hours at 2500°, 15½ hours at 2600°, and ¾ hour at 2700°. The electron projection tube showed only very small crystals and optical microscope examination showed that some crystals were several wire diameters long and occupied the entire cross-section of the wire. These results suggested that a moving temperature gradient similar to that used by the writer on tungsten might extend such crystals to greater length.[10]

For this purpose, an arrangement sketched in Fig. 2 was used.[11] The weight was a piece of tungsten rod 0.060 inch in diameter and ¾ inch long. The mercury was distilled in. The adjustable leak was a Hoke

* This work was started with the aid of a Research Corporation Grant and was completed with the support of a contract with the U. S. Office of Naval Research. It has been reported in more detail in Final Report N6-onr-24433, June, 1952 (unpublished).
† Now at the University of Michigan, Ann Arbor, Michigan.
[1] See, for example, G. F. Smith, preceding paper [Phys. Rev. **93**, 295 (1954)], and references thereto.
[2] C. J. Smithells, *Tungsten* (D. Van Nostrand Company, Inc., New York, 1936), second edition.
[3] C. S. Robinson, J. Appl. Phys. **13**, 647 (1942).
[4] B. A. Mrowca, J. Appl. Phys. **14**, 684 (1943).
[5] By private correspondence with Dr. L. F. Yntema of the Fansteel Corporation.
[6] Johnson, White, and Nelson, Rev. Sci. Instr. **9**, 253 (1938).
[7] L. Malter and D. B. Langmuir, Phys. Rev. **55**, 743 (1939). See also D. B. Langmuir and L. Malter, Phys. Rev. **55**, 748 (1939).
[8] R. P. Johnson, Phys. Rev. **54**, 459 (1938).
[9] Tantalum wire does not have as much tensile strength as tungsten wire so that tungsten spiral springs, which are used extensively with tungsten filaments, are not suitable for tantalum. Worthing (see reference 12) gives data on the thermal expansion of tantalum from which the cold extension of a spring can be calculated so as to let the wire go slack at any desired temperature. However, no data on the tensile strength of tantalum as a function of temperature are available. Flat molybdenum springs in zigzag shape were used and after several tries, the dimensions in Fig. 1(b) were found to give sufficiently weak springs to avoid breaking the wire. The width of the spring was 0.15 cm and the thickness of the molybdenum sheet was 0.003 inch. After forming, the springs were heated in hydrogen to approximately 1200°C for several hours. Springs four cycles in length were used for wires 30 cm long.
[10] M. H. Nichols, Phys. Rev. **59**, 297 (1940).
[11] A nickel wire guide ring, indicated schematically at *A*, was necessary because, without it, surface tension effects in the mercury pushed the weight over to the side of the tube. The position of the guide was such that the surface tension effects applied only between the thin tantalum wire and the guide.

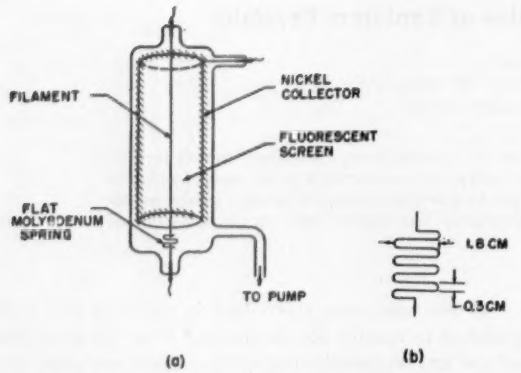

FIG. 1. (a) Schematic diagram of electron projection tube used for observing thermionic emission patterns. (b) Schematic diagram for the zigzag spring.

"OK" needle valve. With this apparatus, four runs were made with cold stretch of 1.2 percent, 2 percent, 2½ percent, and 4 percent. Temperatures from 2200°K to 2600°K were used during the lowering of the mercury. Mercury lowering rates ranged from 4 cm to 0.12 cm per hour. In the case of each wire, the effect of growing "seed" crystals prior to the lowering of the mercury was tried. This was done by a "static" heating schedule, similar to that of the first test run, with the mercury stationary. At the completion of the schedule, the mercury was first raised to overlap the seed crystal, formed just above the mercury during the heating, and then lowered at the desired rate. It was found that seed crystals were not necessary inasmuch as about as many of the longer crystals appeared to be started during the lowering as from seed crystals. From these runs, one crystal 1.5 cm long, one 1.2 cm long, and numerous crystals ranging up to 8 mm in length were obtained. The best results were obtained with a cold stretch of 2 percent, temperature 2400–2500°K, and lowering rate about 0.3 cm/hr. The 1.2 percent stretch produced only very small crystals and the 4 percent stretch appeared to be less effective than the 2 percent in producing larger crystals. This is in agreement with Mrowca's results.[4] Figure 3 is a photograph of a projection tube showing one crystal 1.5 cm long just below level No. 1 and one 1.2-cm crystal just below level No. 4 grown at 2400°K and lowering rate 0.35 cm/hr and 2500°K and lowering rate 0.3 cm/hr, respectively. The 1.2-cm crystal was more perfect than the 1.5-cm crystal.

One run was tried with helium at a pressure of 4.5 cm Hg as a filler gas in order to increase the temperature gradient at the surface of the mercury. The wire was raised to a temperature estimated by the color to be between 2500 and 2600°K and the mercury lowered at a rate of 0.2 cm/hr. After the mercury had gone down approximately 1.2 cm, the wire burned out. Subsequent electron projection tube examination showed no crystals over millimeter length.

The foregoing experiments should not be considered as a complete investigation of the crystal growth properties of tantalum, but rather as a background for further experiments.

II. CRYSTAL GROWTH IN MOLYBDENUM

One run was made on molybdenum wire using the apparatus of Fig. 2. The wire was Fansteel MOLY-G spec. 1.100–17 percent No. R8130-9 dated March 13, 1950. The writer has been informed[5] that the wire "conforms to our specifications which designate limits; namely, iron 0.05 percent max, carbon 0.03 percent max, and molybdenum 99.9 percent min." The data of Worthing[12] on the emissivity and resistivity of molybdenum were used in comparison with the same quantities relative to tungsten[13] to obtain an approximate temperature scale. The mercury was lowered at rates ranging from 0.3 to 1.1 cm per hour at temperatures between 2000 and 2050°K. Figure 4 is a photograph of the thermionic projection tube containing this wire. Because the wire above No. 3 was thinned by evaporation, only the emission above this level shows in the photograph. Approximately ⅔ of the way from level 1 to level 3, several crystals jointly occupied the cross section of the wire but they extended several centimeters along the wire. The seed crystal technique was tried several times in this run but yielded no positive results.

III. THERMIONIC PROJECTION TUBE PATTERNS

The thermionic projection tube patterns of tantalum and molybdenum appeared to be qualitatively identical to that of tungsten.[1] The crystals are oriented with a

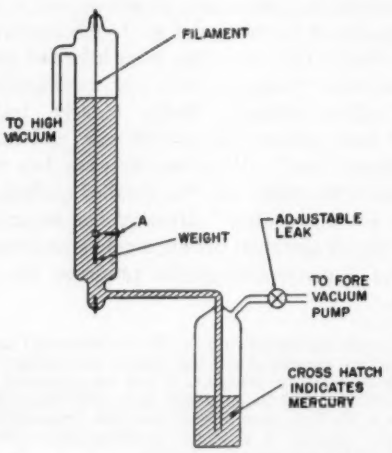

FIG. 2. Schematic of apparatus for producing moving temperature gradient.

[12] A. G. Worthington, Phys. Rev. **28**, 190 (1926).
[13] H. A. Jones and I. Langmuir, Gen. Elec. Rev. **30**, 310, 354, 408 (1927).

(110) crystal axis aligned with the wire axis. This results from slippage of the crystal planes during the drawing process.[2] Thermionic and field emission patterns of tungsten are qualitatively identical.[14–17] No field emission projection tube patterns for tantalum have been published. Benjamin and Jenkins[17] have published a photograph of field emission from molybdenum. If the "lacy" pattern surrounding the (001) direction is assumed to be the result of impurities, than the pattern agrees with that of tungsten. As best could be determined from the molybdenum crystals of Fig. 4, the thermionic and field emission patterns for molybdenum are also identical. To explain the identical field emission and thermionic emission patterns on the basis of internal effects alone—i.e., on the basis of forbidden electronic energy bands—it would be necessary to have gaps in the allowed bands both at the Fermi level and at the surface barrier level for the (110), (112), and (001) directions.[18] That such should occur for the three different metals (still must be checked for the field emission from tantalum) all in the same way is, of itself, very unlikely. The other explanation of anisotropy of emission is variation in the dipole moments of the surface double layers on the various crystal surfaces.[18] Since the three metals all form body-centered cubic crystals, presumably having similar electronic structure, it would be expected that the double layers would be of similar nature. Thus, based on these experimental results alone, it is reasonable to conclude that the anisotropy in the field and thermionic emission patterns of these metals is due mainly to differences in the moments of the double layers rather than to gaps in the allowed energy bands.

FIG. 4. Photograph of electron projection tube containing the molybdenum wire.

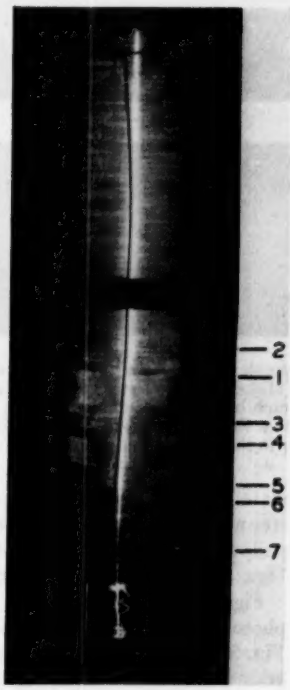

FIG. 3. Photograph of electron projection tube showing thermionic emission pattern from single crystals of tantalum. A mechanical support blocks the view of the tube at about the middle. The nickel wire for collecting the secondary electrons from the fluorescent screen casts a dark shadow; the vertical collector wire is almost coincident with the hot filament.

IV. SURFACE STRUCTURE OF TANTALUM CRYSTALS

The surface structure of the tantalum crystals produced by the techniques described in Sec. I was examined and photographed by the methods described in the accompanying paper.[1] The surface structure of tantalum crystals is similar to that of tungsten in that plateaus are formed normal to certain crystal directions and usually display a shingle-like structure.[1] Figure 5(a) is a photograph of the surface normal to the (110) direction of the lower of the two large crystals appearing in Fig. 3. This was taken after the heating given the wire during the crystal growth procedure, namely 3½ hours at 2500°K and 20 minutes at 2600°K. Careful examination of the other surfaces showed in some places a very faint plateau structure normal to the (001) direction but it was too faint to photograph. According to the

[14] E. W. Müller, Z. Physik **120**, 261, 270 (1943).
[15] R. O. Jenkins, Repts. Progr. Phys. **9**, 177 (1943).
[16] R. Haefer, Z. Physik **116**, 604 (1940).
[17] M. Benjamin and R. O. Jenkins, Proc. Roy. Soc. (London) **A176**, 262 (1940).
[18] C. Herring and M. H. Nichols, Revs. Modern Phys. **21**, 185 (1949), especially pages 203 and 247.

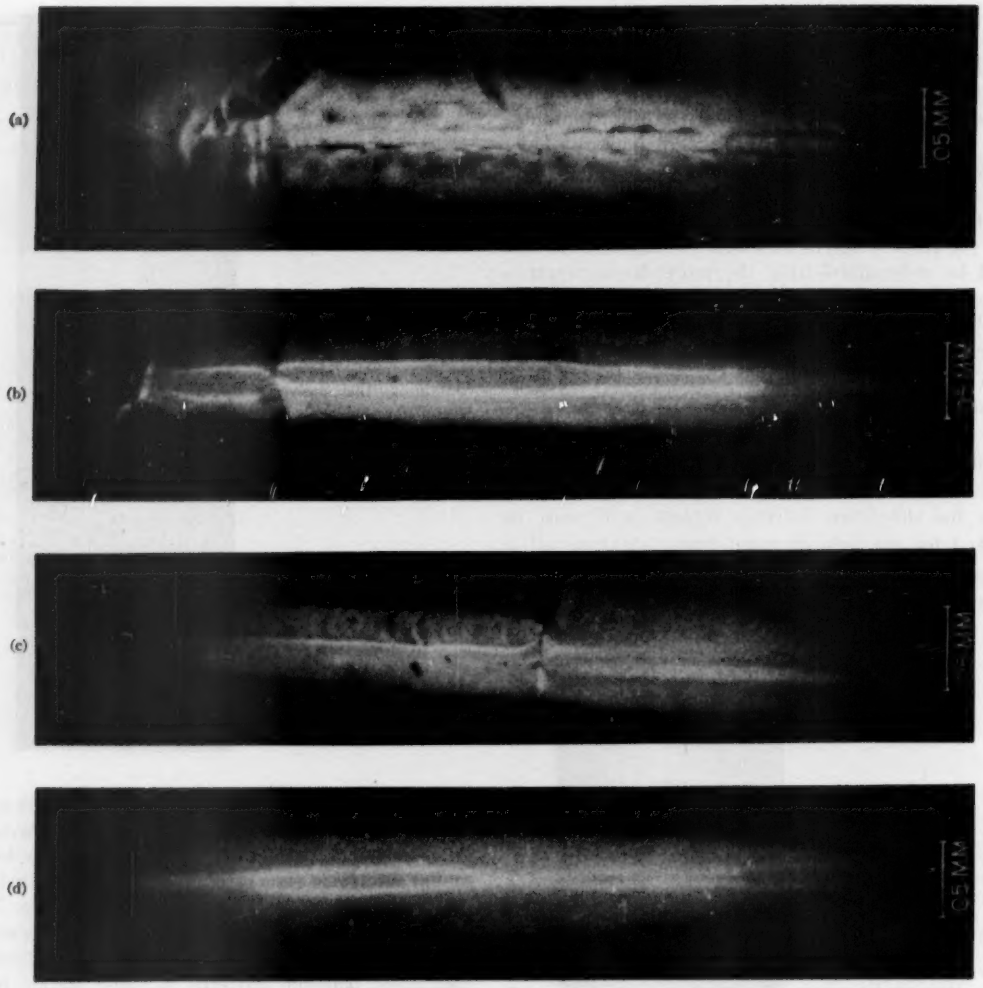

FIG. 5. Photomicrographs of tantalum crystal. (a) Photomicrograph taken along a (110) crystal direction after 3½ hours at 2500°K and 20 minutes at 2600°K. (b) Photomicrograph along (110) direction of same crystal as (a) after evaporating away approximately 5 percent of the wire diameter. (c) Photomicrograph of a grain boundary showing two plateaus. The larger plateau is normal to a (110) direction and the smaller is normal to a (001) direction. (d) Photomicrograph of portion of same crystal as in (b) taken along (110) direction. This shows variation in the plateau width.

data of Langmuir and Malter,[7] the heat treatment up to this point would produce, by evaporation, a reduction in wire radius of 10^{-5} cm.

The section of the wire including the two large crystals of Fig. 3 was mounted in a Pyrex tube containing a barium-aluminum getter bulb. This tube was evacuated, baked, gettered, and sealed off with the usual precautions. Although no ionization gauge was used, a leak test using a mercury cutoff and the system McLeod gauge indicated a tight tube. The wire was then heated according to the schedule of Table I. The temperatures in this table were calculated from the original wire diameter corrected for evaporation using the data of Langmuir and Malter[7] and are only approximate. After approximately 5 percent of the wire diameter was evaporated, the wire burned out. The wire was removed from the tube and the photomicrographs of Figs. 5(b), (c), and (d) were taken.

Figure 5(b), according to the writer's notes, is a photograph of the same section of the wire as shown in Fig. 5(a). It is hard to explain the change in apparent grain boundaries, etc., in Fig. 5(b), but it is certain that both figures are photographs along a (110) direction. Structure could be found normal only to the (110) and (001) directions; the rest of the surface was smooth to within the resolution of the microscope except for the grain boundaries. The (110) plateau of Fig. 5(a) is approximately 0.01 mm wide. In Fig. 5(b), the (110)

plateau has grown to twice this width and the (001) plateau in Fig. 5(c) is approximately 0.01 mm wide. Since the diameter of the wire was 0.12 mm, a width of 0.01 mm subtends an angle of about 10° at the center of the wire. Figure 5(c) shows a crystal boundary and (110) and (001) plateaus for comparison. The wider plateau is the (110).

Figure 5(d) shows a (110) plateau of varying width. To a lesser extent, the width of the (110) plateau of Fig. 5(b) also varies. It is interesting to note that the shingle effect is not discernible in the region of narrowing width but is usually observed in the constant-width plateaus, as in Fig. 5(c), for example. This might imply that the region of narrowing plateaus is one of approach to an ideal crystal plane so that the plateau growth is inhibited.[1] A slight kink in the wire, changing wire diameter, or a slight bending of the crystal could bring this condition about.

In general, the features observed and discussed relative to tungsten in the accompanying paper[1] are also applicable to tantalum. There are two features which appear to be different, however. One is that no plateau was discernible normal to the (112) directions in tantalum. The other is that the heat treatment between the taking of Fig. 5(a) and Figs. 5(b), (c), (d) approximately doubled the width of the (110) and (001) plateaus, whereas in tungsten such heat treatment did not appreciably change the plateau widths. Also, no interferometer measurements were made of wire diameter so that the effect in Fig. 3 of the accompanying paper[1] was not observed.

TABLE I.

Temperature °K	Time at temperature in hours
2260	44
2300	46
2350	48.5
2400	50
2430	46.5
2460	194.5
2400	97.5
2360	48.5
2330	44.3
2360	11

It should be mentioned that a cursory examination of the molybdenum crystals of Fig. 4 also revealed plateau and shingle structure.

Paramagnetic Resonance of Copper Complex Salts

TOSIHIKO OKAMURA AND MUNEYUKI DATE
Research Institute for Scientific Measurements, Sendai, Japan
(Received August 5, 1953)

A paramagnetic resonance experiment was performed on $Cu(NH_3)_4SO_4\cdot H_2O$ and $Cu(NH_3)_4(NO_3)_2$. The character of the $Cu(NH_3)_4^{++}$ group was analyzed and it was found that the g values of this group are small compared with that of the $Cu(H_2O)_4^{++}$ group. This can be explained by two methods, one of which is the covalent bonding hypothesis, the other being the usual ionic hypothesis. The exchange interaction of these salts is very large, which is considered to be the result of covalent bonding structure.

FROM the results of many experiments already performed by many investigators, it has been made clear that the ground state of the paramagnetic Cu^{++} ion in crystals is ^2D, and because of perturbations such as the crystalline electric field and spin-orbit coupling, the effective g value obtained in the paramagnetic resonance experiment varies from about 2.05 to 2.40 when the angle between the crystal axis and the external magnetic field is varied. These experiments were performed satisfactorily on some hydrated crystals such as $CuSO_4\cdot 5H_2O$, copper Tutton salts, etc.[1] Recently, Bleaney and Ingram studied the paramagnetic resonance of ferric complex salts,[2] which have been considered as typical examples of materials having anomalously low magnetic moments. As already pointed out by Abragam and Pryce,[3] the hypothesis of a covalent bond which might exit between the Cu^{++} ion and the surrounding H_2O molecules is unsuitable for these crystals. Thus, the Cu^{++} ion and the surrounding H_2O are only in a coordinated configuration and do not form a covalent bond structure.

On the other hand, it has been pointed out by Pauling[4] that the $Cu(NH_3)_4^{++}$ group forms a covalent bond of $(3d)(4s)(4p)^2$ structure. We measured the g values of this complex group, using single crystals of the two salts $Cu(NH_3)_4SO_4\cdot H_2O$ and $Cu(NH_3)_4(NO_3)_2$ at room temperature, by means of paramagnetic resonance absorption at 21 000 Mc/sec. Figure 1 shows the absorption curves of powdered ammonium complex salts, and also that of powdered copper sulfate for comparison with the former. From this, it is obvious that the g values of the complex salts are quite close to 2. Figure 2 shows the values of g as a function of the angles between the crystal axes and the external magnetic field in a $Cu(NH_3)_4SO_4\cdot H_2O$ single crystal. From these data, we can get the principal values of the g tensor as shown in Table I.

Unfortunately, we cannot deduce the values of $g_\parallel$ and $g_\perp$ because of the lack of x-ray data, but comparing with the values of $g_\alpha=2.25$, $g_\beta=2.20$, and $g_\gamma=2.05$ obtained in the case of $CuSO_4\cdot 5H_2O$, the values of $g_\parallel$ and $g_\perp$ in our case may be expected to be comparatively small and near to 2. To explain this result, we performed the following calculation. As has been already mentioned, the electronic structure of $Cu(NH_3)_4^{++}$ is $(3d)(4s)(4p)^2$. It is easily deduced that one $4p$ electron is almost free, and perhaps its state may be $4p_z$ because of the covalent bond structure.[3] However to calculate the deviation of the g values from 2, we started with

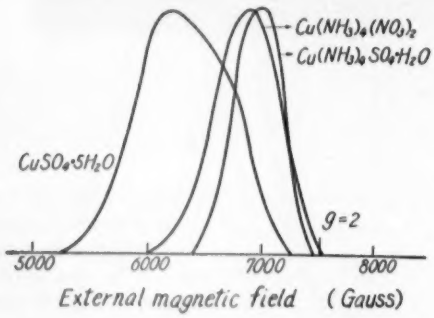

FIG. 1. Resonance curves of copper ammonium complex salt (powder).

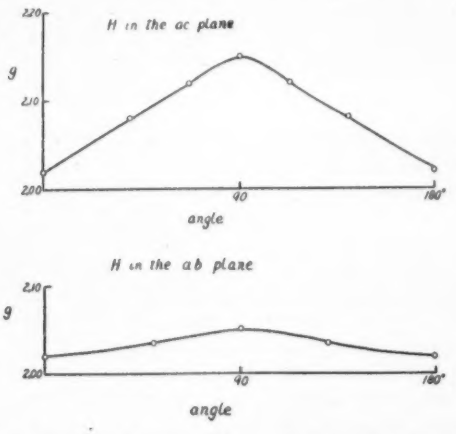

FIG. 2. Variation of g values with the angle between the static field H and the crystal axis in $Cu(NH_3)_4SO_4\cdot H_2O$.

[1] For example, R. D. Arnold and A. F. Kip, Phys. Rev. **75**, 1199 (1949).
[2] B. Bleaney and D. J. E. Ingram, Proc. Phys. Soc. (London) **A65**, 953 (1952).
[3] A. Abragam and M. H. L. Pryce, Proc. Roy. Soc. (London) **206**, 164 (1951); A. Abragam, Physica **17**, 209 (1951).
[4] L. Pauling, *Nature of the Chemical Bond* (Cornell University Press, Ithaca, 1940).

TABLE I. The g values of copper ammonium complex salts.

Substance	g_x	g_y	g_z
$Cu(NH_3)_4SO_4\cdot H_2O$	2.02±0.02	2.05±0.02	2.15±0.02
$Cu(NH_3)_4(NO_3)_2$	2.07±0.02	2.14±0.02	2.02±0.02

^2P as the ground state of the ion, and considered that the contribution of other higher states is negligible. We then introduced the crystalline electric field, the spin-orbit interaction, and the external magnetic field as perturbations.[5]

As the Cu^{++} ion is placed at the center of four NH_3 molecules which are in a plane, we may assume the symmetry of the electric field to be tetragonal, so that the Hamiltonian is given by the following equation:

$$H=H_0+D(x^4+y^4+z^4)+A(x^2+y^2)-2AZ^2 +\lambda(\mathbf{L}\cdot\mathbf{S})+\beta\mathbf{H}\cdot(\mathbf{L}+2\mathbf{S}).$$

From the above equation, we get the following results:

(1) The ^2P state is not split by the spherically symmetric field.

(2) The degenerate level is split by the tetragonal field into a singlet and a doublet and this splitting is $(6/5)A\langle\gamma^2\rangle_{Av}$.

(3) In addition, because of the $(\mathbf{L}\cdot\mathbf{S})$ coupling, the doublet state is separated by an amount of the order of λ. These results are shown in Fig. 3, and the energies are expressed as follows:

$$E_0\longrightarrow W_0=\tfrac{1}{2}\{E_0+E_1-\tfrac{1}{2}\lambda+[(E_0-E_1-\tfrac{1}{2}\lambda)^2-2\lambda^2]^{\frac{1}{2}}\},$$
$$E_1\begin{cases}\nearrow W_1=E_1+\tfrac{1}{2}\lambda,\\ \searrow W_2=\tfrac{1}{2}\{E_0+E_1-\tfrac{1}{2}\lambda-[(E_0-E_1-\tfrac{1}{2}\lambda)^2-2\lambda^2]^{\frac{1}{2}}\}.\end{cases}$$

(4) The effective g values for each level when the external magnetic field is applied are calculated and are shown in Table II.

From the above results, we can conclude that the lowest state is W_0 (using the notation of Table II). The value $g_{11}=2$ in the case of the $Cu(NH_3)_4^{++}$ group

TABLE II. The g values of each level.

level	g_{11}	$g_{\perp}$
W_0	2.00	$2\left(1-\dfrac{\lambda}{E_0-E_1}\right)$
W_1	4.00	0
W_2	0	$\left\lvert\dfrac{2\lambda}{E_0-E_1}\right\rvert$

[5] D. Polder, Physica 9, 709 (1942).

may be evidence of covalent bond structure, in agreement with Abragam's work[3] and the present experimental result. Moreover, it is noticed that the sign of the tetragonal field constant A is positive while it is negative in the case of the usual Cu^{++} ion, and the values of the splitting due to the tetragonal field can be calculated to be about 6000 cm^{-1} from our experimental data if we assume that the value of λ is equal to that for the Cu^{++} ion. But this value may be incorrect. Because of the covalent bond structure, the ^2P state is already affected before we consider the crystalline electric field. So in our calculation this effect must be considered, and this is accomplished by varying the $\mathbf{L}\cdot\mathbf{S}$ coupling constant λ.

In addition to the above considerations, we must consider the fact that these experimental results can also be explained by the usual method calculated by Polder if we assume the crystalline cubic electric field to be two or three times larger than that of ordinary paramagnetic copper salts. Because of the lack of x-ray data, we cannot say whether the presence of this large electric field is probable. So, as the result of these

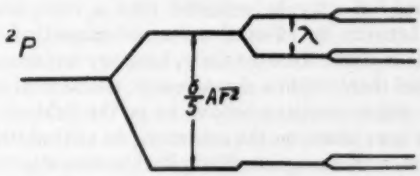

FIG. 3. Energy-level splitting of ^2P state.

considerations, we must consider at least two possibilities for the ground state of the $Cu(NH_3)_4^{++}$ group as distinct from the $Cu(H_2O)_4^{++}$ group.

Another interesting result is the half-width of the absorption lines. Because of the lack of x-ray data, a rigorous application of Van Vleck's theory of dipolar broadening[6] is impossible; however, using the density of the single crystal, we can calculate from the dipolar interaction an approximate value of the half-width of about $\Delta H\sim 400$ gauss. But the experimental data show about 30~50 gauss in various directions in the case of $Cu(NH_3)_4SO_4\cdot H_2O$, and also a slightly larger value for the nitrate. Thus, our experimental results give about $\frac{1}{10}$ of theoretical values. Hence, very strong exchange interactions may be expected. This may be caused by the effect of a larger mean distance for the $4p$ orbit than for the $3d$ orbit in the case of Cu^{++} exchange interactions in the coordinated configuration.

The authors are indebted to Professor M. Kotani at Tokyo University for helpful discussions on the problem of complex ions.

[6] J. H. Van Vleck, Phys. Rev. 74, 1168 (1948).

Behavior of Ferromagnetic Domains under Stress

G. BONFIGLIOLI, A. FERRO,* G. MONTALENTI, AND G. ROSA
Istituto Elettrotecnico Nazionale, Turin, Italy

(Received August 3, 1953; revised manuscript received December 1, 1954)

The results of some experiments referring to the analogy between the effects of H (uniform magnetic field) and σ (static mechanical stress) on the motion of Weiss domains are described. Together with the results of previous work, the present paper attempts to show the existence of a magnetoelastic Barkhausen effect. A special extensometer with a strain sensitivity of 10^{-7} and completely free from hysteresis was used. Experiments have also been carried out on electrical resistivity variation. The results are summarized as follows: the curves of the remanent strain, $\epsilon_r=(\Delta l/l)_r$ *versus* σ_{max}, and of the remanent electrical resistivity variation, $(\Delta\rho/\rho)_r$ *versus* σ_{max}, traced for two Ni specimens, one as cast and one annealed, show the same behavior as J_r (remanence magnetization) *versus* H_{max}. Both remanent strain and remanent electrical resistivity variation can be reduced to zero by demagnetizing the specimen. By means of these experiments the characteristic points of the magnetomechanical hysteresis loop have been determined statically and compared with the magnetic case.

1. INTRODUCTION

MOST of the effects of stress on ferromagnetic properties have already been studied by different authors,[1] but it is still worth while to devote some attention to certain aspects of this problem.

As a result of previous work by some of the present authors,[2,3] it may be assumed that a close analogy holds between the effect of stress and magnetic field on domain motion. More precisely, for every ferromagnetic material there exists a characteristic mechanical stress, below which domains behave as in the field of Rayleigh's law; when, on the contrary, the critical value is reached, a true magnetoelastic Barkhausen effect takes place, with irreversible domain jumps followed by a quasi-elastic rotation. The analogy is very close, the only difference being that in the magnetoelastic case the macroscopic intensity of magnetization must obviously remain zero and the orientation of domain magnetization *versus* stress will depend on the sign of the magnetostrictive coefficient, being parallel or perpendicular according to its sign.

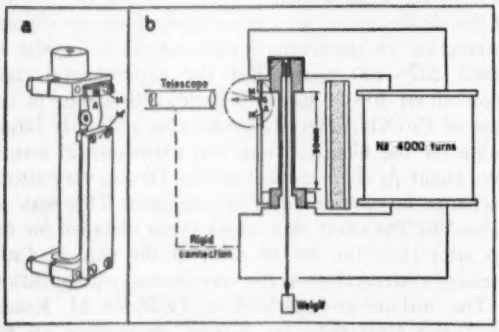

FIG. 1. (a) The specimen with the extensometer mounted on it. (b) Schematic view of the apparatus used.

It should be noted that the magnetoelastic Barkhausen effect has been directly observed by a method strictly analogous to the one for the magnetic case, and described in a paper[4] which was unknown to the writers when their previous work was published. The abovementioned analogy suggests the existence of magnetomechanical energy losses corresponding to those of magnetic hysteresis.

An examination of the literature on the problem of energy losses in mechanical loops of ferromagnetic materials shows that only the field of low stresses has been thoroughly investigated. For the dynamic case, the existence of losses proportional to the frequency of vibration has been shown. For the static case, there exist some papers[5,6] on the "remanent strain" but still in the field of low stress.

As to the results of references 2 and 3, it has been shown that: (a) all the curves of J_r/J_s *versus* σ, δ *versus* σ, and J *versus* H show the same behavior (J=intensity of magnetization, J_s=saturation magnetization, J_r=remanence magnetization, H=magnetic field, σ=mechanical stress, δ=energy losses, J_r/J_s =parameter related to domain distribution); (b) a law similar to the Steinmetz empirical law holds also for the magnetomechanical energy losses; (c) magnetic hysteresis energy losses in a saturation loop are of the same order as magnetoelastic ones (in a mechanical cycle described with a very low frequency, between values of stress not far from the yield point). These measurements have been made with a torsion pendulum.

2. PURPOSE OF PRESENT INVESTIGATION

It seemed that it should be of some interest to investigate these phenomena in the static case, for homogeneous deformation and for stresses high enough to give rise to irreversible domain jumps. Because of experimental difficulties it was impossible to draw directly

* Fiat Central Laboratory, Turin, Italy.
[1] R. M. Bozorth, *Ferromagnetism* (D. Van Nostrand Co. Inc., New York, 1951), p. 595 ff.
[2] A. Ferro and G. Montalenti, J. Appl. Phys. **22**, 565 (1951).
[3] Bonfiglioli, Ferro, and Montalenti, Phys. Rev. **86**, 959 (1952).
[4] Toru Kamei, J. Phys. Soc. Japan **6**, 260 (1951).
[5] G. Richter, *Probleme der Technische Magnetisierungskurve* (Springer, Berlin, 1938), pp. 93–113.
[6] W. Döring, Ann. Physik **32**, 259 (1938).

the magnetomechanical hysteresis loop, as would have been desirable. The scope of the present work is therefore limited to the following points: (a) to show that the remanent strain *versus* σ_{max} after release of a static mechanical stress in pure extension has the same behavior as a magnetization curve; (b) to show that this remanent strain may be reduced to zero by demagnetizing the specimen (by successive hysteresis loops tending to zero): only in this way can we be sure that there are no errors due to plastic deformation; (c) to show that a similar irreversible process exists also with the electrical resistivity.

3. DESCRIPTION OF APPARATUS

The equipment used consists of a static test machine, with some particular features to obtain high accuracy in the centering of the load on the specimen by wire suspension, perfect load reproducibility by direct load application, absence of shocks during the charge with an hydraulic device, possibility of applying a magnetic field to the specimen, thermal screening between field

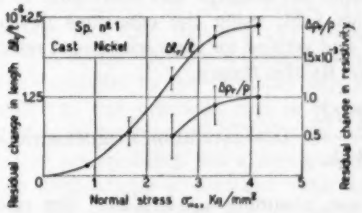

FIG. 2. Curves of remanent strain *versus* maximum stress and of remanent resistivity variation *versus* maximum stress for cast nickel. The uncertainty in the ordinate scale of $\Delta l_r/l$ about ± 40 percent.

wiring and specimen in order to avoid thermal drift, etc. The applied loads are of compression according to the fact that magnetostriction of the specimen is negative.

The extensometer for the strain measurements is of a new type specially designed in order to obtain the necessary very high sensitivity and absolute absence of hysteresis. Since remanent strain is of the order of 10^{-6} a strain sensitivity of 10^{-7} is required. Some initial attempts with other methods such as light fringes or strain gauges were unsuccessful. The main difficulty was to obtain an extensometer completely free from elastic hysteresis even after a large deformation, since the problem was the precise detection of the very little remanent strain in the specimen after the removal of the load. The extensometer shown in Fig. 1(a) consists of a plane phosphor bronze spring A, rigidly clamped to points $M-M'$ and elastically predeformed just beyond the point of flexural instability. A mirror G is rigidly fixed at one fourth of the distance MM', and a light beam reflected on G rotates for an appreciable space when distance MM' is slightly changed. An optical lever with a high-power telescope (equivalent lever

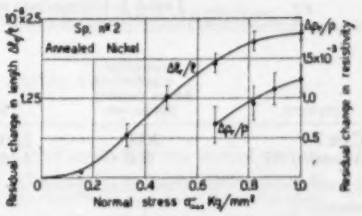

FIG. 3. Curves of remanent strain *versus* maximum stress and of remanent resistivity variation *versus* maximum stress for annealed cast nickel. The uncertainty in the ordinate scale of $\Delta l_r/l$ is about ± 40 percent.

length 10 m) allows a sensitivity of about 0.02μ per unit scale.

From measurements on a duraluminum specimen, we observed that this device is entirely free from hysteresis, as stress at A is negligible and all connections are completely rigid. It is, however, evident that such a device is not linear, but calibration may easily overcome such a difficulty. Calibration using as reference the elastic modulus of the specimen must be repeated every time the extensometer is removed from the specimen, as the sensitivity is greatly influenced by the initial MM' distance. Another possible cause of error could have been a defect in axiality of load or homogeneity of material: however, a series of readings taken with the extensometer at 0° and 180° azimuth showed differences always less than 10 percent and therefore not significant to our grade of accuracy. The thermal screening between specimen and field wiring did not give any detectable drift even for periods of time much longer than that required for demagnetizing the specimen.

From the examination of root-mean-square errors of the many series of readings taken during the experiments, we conclude that the reproducibility of measurements within a series of readings is better than ± 10 percent; and that the uncertainty in the absolute calibration of the extensometer is restricted to ± 40 percent. Results may nevertheless be considered satisfactory and allow sufficient reliability as to the description of the phenomenon.

The drawing of Fig. 1(b) shows schematically the equipment used.

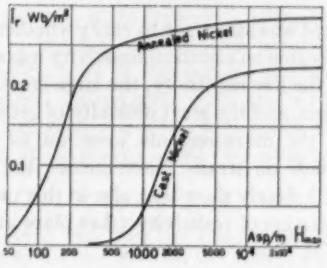

FIG. 4. Curves of remanent magnetization (in webers/meter²) *versus* maximum field (in ampere turns/meter) for cast nickel and annealed nickel.

TABLE I. Comparison of static and dynamic magnetomechanical losses.

Specimen	Saturation magnetic losses (static) joules/m³	$10^4 \cdot \epsilon_r$	σ_{max} kg/mm² (sat. value)	Magnetomechanical losses (static) joules/m³	σ_{max} kg/mm² (sat. value)[a]	Magneto-mechanical losses (dynamic $4H_s$) joules/m³ [a]
Cast Ni	2000	2.30±40%	4.25	104÷416	4	480
Annealed Ni	700	2.25±40%	1	24÷ 96	1	180

[a] From reference 3.

4. RESULTS

The results of the measurements are shown in Figs. 2 and 3. The curves give the remanent strain $\epsilon_r = (\Delta l/l)_r$ and the remanent percent resistivity variation $(\Delta\rho/\rho)_r$ *versus* applied maximum stress. Curves were obtained on Ni specimens: in Fig. 2 for cast Ni and in Fig. 3 for the same Ni annealed. The coercive forces of the two specimens were, respectively, 1200 Asp/m (ampere turns per meter) for the cast nickel and 200 Asp/m for the annealed one. Domain distributions in the specimen are nearly identical and probably nearly isotropic: $J_r/J_s = 0.40$ for specimen No. 1 (cast) and $J_r/J_s = 0.50$ for specimen No. 2 (annealed).

Measurements have been made according to the following scheme: (1) specimen demagnetized (by successive hysteresis loops tending to zero)—unloaded—reading, (2) specimen loaded (till σ_{max}), (3) specimen unloaded—reading, (4) specimen redemagnetized—reading. The length of the specimen in state 4 is practically identical to that in state (1). Taking σ_{max} as abscissa, the corresponding ordinate ϵ_r is given by the difference in length of the specimen between state (1) [or (4)] and state (3). Every experimental point represents an average of 80 readings.

Both curves of remanent strain ϵ_r show a variation of slope for a certain critical stress and for higher values approach saturation: this behavior is strictly analogous to that of the remanent magnetization intensity J_r *versus* maximum field H_{max} and confirms the presence of the magnetoelastic Barkhausen effect (see also the papers quoted[2,3]). (See Fig. 4.) Moreover, in the course of experiments it has been found that the maximum magnetoelastic remanent strain ϵ_r is about $\frac{1}{3}$ of the remanent magnetostrictive deformation measured after the material has been magnetically saturated and then released.

An attempt was also made to verify whether the effect of domain motion on electrical resistivity was analogous. Because of the low resistivity, the large cross section of the specimens, and the great difficulty of getting a good sensitivity, the measurements were not as precise as the ones made on strain. Nevertheless, the curves of Figs. 2 and 3 clearly show that also in this case an irreversible variation of resistivity takes place at a certain critical stress and seems to show a similar behavior. Also, this variation can be reduced to zero by demagnetizing the specimen, and yields about $\frac{1}{2}$ of the remanent magnetoresistivity. We are indebted to Professor A. Drigo for the remark that these results on the electrical resistivity variation seem to give support to the fact that 90° jumps are the principal cause of the effect.

5. EQUIVALENCE BETWEEN FIELD AND STRESS AND COMPARISON WITH MAGNETIC LOSSES

The ratio between coercive forces of the two specimens is nearly identical to the ratio between critical stresses, and the two quantities seem thus to be simply proportional. An attempt was also made to correlate stress with field, using the equations developed by Bozorth[7] and verified for the variations of resistivity on permalloy. By this theory,

$$\sigma = -\frac{2HJ_s}{3\lambda_s} \quad (\lambda_s = \text{saturation magnetostriction}).$$

In our case, assuming $\lambda_s = 34 \times 10^{-6}$ (see reference 1, p. 659), we have $\sigma = 1$ kg/mm² and therefore $H = 900$ ampere turns/meter, and characteristic points of the curves of ϵ_r *versus* stress should correspond to those of curves of J_r *versus* field: the agreement obtained with these data seems fairly good, especially for cast nickel.

From the values of ϵ_r a rough comparison between static magnetomechanical losses with the dynamic ones can also be made. The order of magnitude of the area of the magnetomechanical loop, by analogy with the corresponding magnetic case for the same material, can be assumed equal to $4 \times \frac{3}{4} \times \epsilon_r \times \sigma_{max}$: the approximation with which this value is known may be about ±60 percent; the difficulty of measurement did not allow more accurate values. It is nevertheless possible to confirm the results of a preceding work[3] obtained with damping measurements and to confirm that the magnetomechanical energy losses determined statically or at very low frequency for a stress high enough to obtain magnetomechanical saturation are of the same order of magnitude as the magnetic losses in the saturation loop, the former being nevertheless decidedly inferior (Table I).

[7] R. M. Bozorth, Phys. Rev. **70**, 923 (1946).

Magnetic Susceptibility of Carbons and Polycrystalline Graphites. I*

H. T. PINNICK
Department of Physics, University of Buffalo, Buffalo, New York
(Received September 25, 1953)

In this investigation, the relation of susceptibility to the a dimension of crystallites was studied over the range from turbostratic carbons with diameters of 50 angstroms, to graphitic crystals with diameters of at least several thousand angstroms. Two series of carbons, one Thermax carbon black in powder form, the other National Carbon baked rods, were prepared in which the crystallite size was varied by heat treatment to different temperatures. A third series consisting of carbon blacks with different particle sizes from 50A to 3000A was prepared by heating to 3000°C, the particle size limiting the crystallite growth. The crystallite diameter was determined from the broadening of the x-ray diffraction lines, 10 (two-dimensional asymmetric diffraction line), and 110. The magnetic susceptibility was measured at room temperature by the Gouy method, and the measurements were made with fields ranging from 2000 to 11 500 oersteds to correct for ferromagnetic impurities. The mass susceptibilities of all carbons plotted against crystallite diameter fall within a narrow band showing that the susceptibility is primarily a function of crystallite size (Fig. 5). Almost all the increase in susceptibility from -0.5×10^{-6}/g for very small crystallites (aromatic molecules) to -7.8×10^{-6}/g for large crystals takes place in the range of diameters from 50A to 150A. The sharp increase in susceptibility is discussed in relation to the changes in the Hall effect and thermoelectric power as a function of heat treatment. It is concluded that this increase is due to an increase in the filling of the π band which brings the Fermi level into a region of higher curvature of the constant energy surfaces.

INTRODUCTION

THE susceptibility of large single crystals of graphite was first measured by Ganguli and Krishnan[1] who found that the difference between the susceptibility parallel to the principal axis and that perpendicular to the principal axis is about -21.5×10^{-6}/g at room temperature. This large anisotropic component of susceptibility has a strong temperature dependence approaching a constant value of -30.0×10^{-6}/g at low temperatures and changing into an approximate linear relation with $1/T$ at high temperatures ($T>300$°K). Several investigators have found that when graphite crystals are ground to a size of one micron and smaller the susceptibility becomes considerably smaller.[2,3] For crystallites containing only a few benzene rings the susceptibility should approach a value between -0.5×10^{-6}/g, the value for a carbon atom, and -0.7×10^{-6}/g, the average susceptibility for a randomly arranged system of benzene rings. Since all the change in susceptibility occurs in the region between very small and very large crystals, this is the region in which the relation of susceptibility to size had to be investigated.

Miwa[4] attempted to check the existence of a relation between susceptibility and size from susceptibility measurements on a number of carbon blacks. The blacks used by him had different crystallite sizes resulting from different methods of preparation. His results indicated that the susceptibility increased roughly linearly with the c dimension (thickness of crystallites perpendicular to the graphitic layers). The c dimension of his samples varied from about 10A up to 70A. Since the a dimension is approximately proportional to the c dimension,[5] this would mean that the susceptibility also increases linearly with the a dimension. When the size becomes very large, the susceptibility must approach asymptotically the value for single crystals of graphite. Even if Miwa's results were correct, it is clear that the linear relation can hold only over a limited range of crystallite size. It is generally believed that the large susceptibility of graphite is due to motion of electrons in the planes and that the interaction with neighboring planes is very small. Therefore, it seems that it is the a dimension which should be considered in studying susceptibility. Recently a considerable amount of information on the electronic behavior of carbons as a function of crystallite size has been obtained,[6] and it seemed important to investigate more extensively the relation between susceptibility and crystallite size from hydrocarbons up to graphite in order to check on the proposed explanation.

After this work was well under way, Wynne-Jones, Blayden, and Iley[7] published a study of the susceptibility of carbons with crystallite diameters ranging from 20A to 90A and also of some natural graphites. The crystallite sizes of the carbons studied by them were in general smaller than those studied here. However, the size ranges overlap sufficiently to permit a comparison of the results of the two investigations (see following).

* Supported in part by the U. S. Office of Naval Research.
[1] W. Ganguli and K. S. Krishnan, Proc. Roy. Soc. (London) **117**, 168 (1941).
[2] J. R. Rao, Indian J. Phys. **6**, 241 (1931).
[3] C. H. Gregory, thesis, California Institute of Technology, 1935 (unpublished).
[4] M. Miwa, Sci. Repts. Tôhoku Univ. **23**, 242 (1934).
[5] R. E. Franklin, Proc. Roy. Soc. (London) **209**, 196 (1951).
[6] S. Mrozowski, Phys. Rev. **85**, 609 (1952); Phys. Rev. **86**, 1056 (1952).
[7] Wynne-Jones, Blayden, and Iley, Brennstoff-Chem. **33**, 268 (1952).

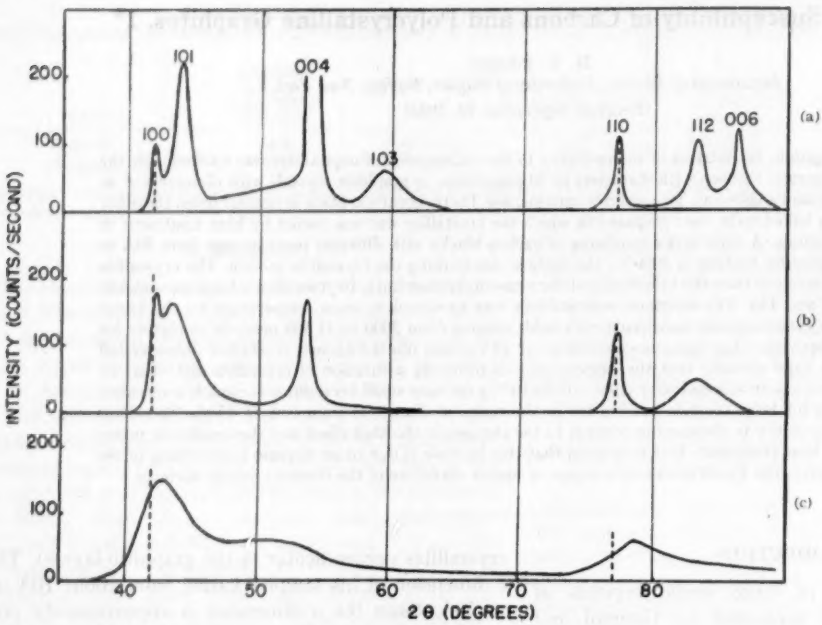

FIG. 1. X-ray patterns of carbons with large graphitic crystals (a), with crystals of intermediate size and layers partially aligned (b), and with small turbostratic crystallites (c).

EXPERIMENTAL AND RESULTS

1. Sample Preparation

Two somewhat different methods were used to obtain various crystallite sizes. The first consisted of heat treating a series of carbon blacks with particle sizes from 50A to 3000A, to a final temperature of 3000°C. It is known from electron microscope studies[8] that in graphitized carbon blacks, the particle size limits the crystallite growth. The series of graphitized carbon blacks, therefore, had different crystallite sizes but all had been heat treated to the same temperature. The second method consisted of heat treating samples to different temperatures to obtain different average crystallite sizes. Two groups of samples were treated in this way, one of commercial baked carbon rods from National Carbon Company, the other a carbon black with a large particle size d (Thermax, $d\sim3000$A). All the samples were heat treated in a furnace consisting of a graphite tube serving as a resistance element and the temperature was measured with a Leeds and Northrup optical pyrometer. An atmosphere of nitrogen was provided in normal runs, but chlorine was also introduced during some of the runs at temperatures around 1000°C to reduce the amount of ferromagnetic impurity.

2. Measurement of Crystallite Diameters

The crystallite sizes of all the carbons were determined from the broadening of x-ray diffraction lines. In the initial stages of the investigation, the x-ray patterns were recorded photographically by the Debye powder method, and the broadening was measured from microdensitometer traces of the films, while later a North American Philips spectrometer was used. For the Debye powder patterns Cu x-radiation was used with a Ni filter consisting of two foils of 0.005-in. thickness placed around the circumference of the powder camera. Uniform size of samples was insured by extruding them through a 0.5-mm capillary. The characteristics of each film were determined with the aid of a rotating sector disk which gave five exposures each differing by a factor of two. These exposures were adjusted to cover the range of densities on most films and except for densities lower than any background, the density $[\log(I/I_0)$ as measured on a Leeds and Northrup microdensitometer$]$ was linearly related to exposure time. Since the density above the background is directly proportional to the exposure, the half widths of the lines were determined directly from the densitometer trace of the total scattering after the background had been subtracted. For samples investigated with the spectrometer, the lines were measured by a fixed count technique to reduce statistical error in the background.

In order to get the broadening β of the line due to the finite size of the crystallites, the instrument width b was subtracted from the total half-width B. The instrument width was obtained in both the Debye and spectrometer methods from the lines of coke graphitized to 3000°C, and of natural graphite in which the crystallites are so large that the size broadenings are negligible. The x-ray pattern of this coke is shown in Fig. 1, along with patterns of a coke of intermediate size, and one of very small size. The relation between broadening of a line and the crystallite dimensions is: $l=C\lambda/\beta\cos\theta$,

[8] H. T. Pinnick, J. Chem. Phys. 20, 756 (1952).

where t is the dimension of the crystal perpendicular to the set of planes giving rise to the line, λ is the wavelength of x-radiation, and C is a constant ordinarily taken to be equal to one. However, small crystallites of carbon have a turbostratic structure with successive planes parallel, but randomly rotated about the direction perpendicular to the planes so that lines corresponding to cross lattice reflections *hkl* of graphite are not present, and only lines corresponding to *hk*0 and 00*l* are observed (Fig. 1c). Warren[9] has shown that the diffuse reflections corresponding to the *hk*0 lines in graphite are the result of two-dimensional diffraction from each plane. He showed that the shape is asymmetric with the intensity falling off more slowly on the large angle side of the peak, and that the constant C in the above relation between broadening and size is equal to 1.85 for two-dimensional diffraction lines. As the crystallite size increases, the layers gradually become oriented in the crystallites in the same manner as they are in the graphitic structure. Franklin[10] has shown that in heat-treated cokes the structure of a crystallite consists of a stack of parallel planes composed of a random mixture of groups of perfectly oriented planes and groups of completely disoriented planes. As the crystallite size increases, the percentage of oriented planes increases and *hkl* lines gradually begin to appear in the x-ray patterns. For example, the 101 line becomes noticeable when it is about $\frac{1}{10}$ as intense as the neighboring 100 line. Experimentally, the beginning of alignment is observed in carbons with quite different histories of temperature treatment in about the same range of sizes, that is, around 150A.[8] Therefore, the orientation of the layers is not caused just by the increased thermal motion at higher temperatures of heat treatment, and probably the orientation process is the result of two competing forces, one tending to align the layers and the other preventing alignment.[8] For crystallites of this order the 101 line shows up on the large angle side of the 100 line in the form of a hump which gives an error of about 20 percent in the determination of crystallite size. Therefore, the 100 line was used only for size measurements of crystallites with diameters less than 150A and above 150A where the alignment gives rise to *hkl* lines, the size was obtained from the half-width of the 110 line which is sufficiently intense at these sizes for accurate half-width determination (Fig. 1b).

3. Magnetic Susceptibilities

The magnetic susceptibility was measured, using a Gouy balance. A regular chainomatic analytical balance was modified so that the sample could be placed in the field of an electromagnet by directly suspending it from the balance arm. The sample and the nylon suspension were enclosed by a glass tube to protect them from air currents. The separation of the pole pieces was approximately 2.0 cm and at this separation the maximum field obtainable was 11 500 oersteds, sufficient to achieve saturation of ferromagnetic impurities. The field intensities were obtained in each case by measuring the current in the coils, the measurement being made with a precision ammeter ($\frac{1}{2}$ percent). A calibration curve of field intensity *vs* current was obtained from fifteen determinations of the field made with a fluxmeter at current settings approximately the same as those used in the susceptibility measurements, and this curve was used to interpolate for fields corresponding to intermediate values of current. A special search coil was used for most of the higher field strengths and it was calibrated with the standard search coil supplied with the instrument. No current regulating device was employed, since the current through the magnet was constant to within 1 percent over the period required for determining the force on the sample, and such a change in current causes an even smaller change in the field.

The susceptibility of a powder sample was measured by placing the powder in the upper part of a special container[11] divided in the center by a flat glass baffle, with the lower half of the container evacuated. The samples which were in the form of rods were suspended so that the lower end of the rod was in the center of the region of uniform field. It was necessary to consider the correction for oxygen in the air displaced by the rod. For the atmospheric concentration of 20 percent this correction has the value $\kappa=+0.03\times10^{-6}$/cc, which was negligible for all samples in this investigation.

The volume susceptibility of all samples, both powders and rods, was measured at room temperature for 15 different field strengths ranging from 2000 oersteds to 11 500 oersteds. The experimental values of the volume susceptibilities were plotted against $1/H$ and this curve extrapolated to $1/H=0$ to correct for ferromagnetic impurities. This procedure does not correct for any paramagnetic impurities, however.

In most instances the correction for ferromagnetic impurities ($\kappa_{5000}-\kappa_{\infty}$) amounted to less than 20 percent. Commercial graphite rods contain relatively large amounts of residual ferromagnetic material from the Fe_2O_3 added to them to inhibit puffing during graphitization and one of these rods was used to check the validity of the procedure. The susceptibility of this rod was measured, and the extrapolated susceptibility obtained was -12.7×10^{-6}/cc, a correction ($\kappa_{5000}-\kappa_{\infty}$) of 100 percent (see curve *a*, Fig. 2). Following LaRosa[12] the rod was purified by heating it in a chlorine atmosphere at 250°C for 30 minutes and the curve of κ *vs* $1/H$ after this treatment is shown in Fig. 2b. Finally, the rod was heated in a chlorine atmosphere at 900°C for 30 minutes and the resulting curve of κ *vs* $1/H$ is shown in Fig. 2c. The larger diamagnetic susceptibility in Fig. 2c at low fields, as well as the smaller value of the

[9] B. E. Warren, Phys. Rev. **59**, 693 (1941).
[10] R. E. Franklin, Acta Cryst. **4**, 254 (1951).
[11] P. W. Selwood, *Magnetochemistry* (Interscience Publishers, Inc., New York, 1943).
[12] M. LaRosa, Nuovo cimento **12**, 284 (1916).

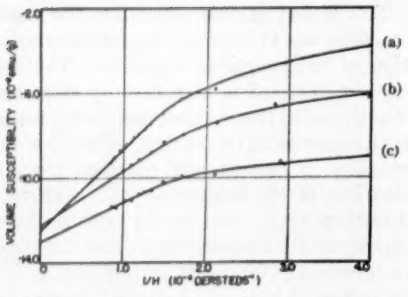

FIG. 2. Volume susceptibility as a function of the reciprocal of the magnetic field intensity for a graphite rod with ferromagnetic impurities removed by successive heat treatments in an atmosphere of chlorine: (a) untreated; (b) treated at 250°C; (c) treated at 900°C. *Correction added in proof:* units are (10^{-6} emu/cc).

slope of κ *vs* $1/H$, shows a considerable reduction of the amount of ferromagnetic material. However, the extrapolated value of κ obtained from the curve, -13.1×10^{-6}/cc, differs from the extrapolated value for the untreated rod by only 2 percent. Since this treatment would probably have removed a large part of any paramagnetic impurities, the very small change shows that the amount of paramagnetic impurities is negligible. It also makes it very probable that the extrapolated values are quite reliable in the other cases where the correction ($\kappa_{1000} - \kappa_{\infty}$) was only 20 percent or less.

In the measurement of susceptibility of powders by the Gouy method, the uniformity of packing is the main factor limiting the accuracy of the final results. The column of powder in the container described above was about 16 cm long, and to obtain uniform packing the powder was loaded in 10–12 consecutive operations, each one filling about 1.5 cm of the tube. The data obtained for the packing of a typical powder sample (Fig. 3) show that the density of the powder column is constant to within ± 2 percent. A further check on the uniformity of packing was made by measuring the susceptibility with the container suspended at one position in the field, and then raising the container by half the average length of one of the sections in which the powder was packed and again measuring the susceptibility. For all samples the results were identical to within ± 1 percent.

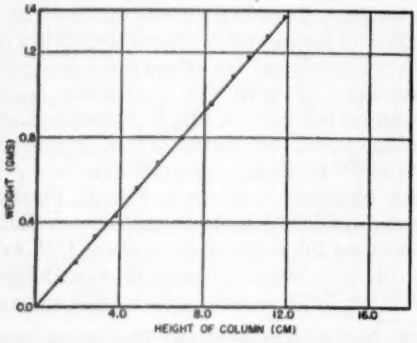

FIG. 3. Density of a carbon black sample as a function of height of column.

4. Results

The results of crystallite size measurements and susceptibility determinations are given in Table I.

There are a number of points in Table I which require an explanation. In the case of powders, the tamping produces orientation with planes perpendicular to the axis of the sample tube and consequently decreases the measured value for the susceptibility since in this direction the susceptibility is -0.5×10^{-6}/g for all sizes. For example, two powder samples of natural graphite were measured without tamping, and the values obtained for the susceptibility were about -6×10^{-6}/g. For these samples the holder was tapped to pack the powders, and even without tamping the plate like structure of the particles with the c axis of the crystals perpendicular to the plates resulted in a preferred orientation of crystal planes perpendicular to the axis of the sample tube. From electron micrographs of shadowed replicas of graphitized Thermax, it has been established[13] that the structure of these carbon blacks changes during graphitization from spherical into polyhedra with the crystallite planes parallel to the faces of the polyhedra. Therefore, when the carbon blacks are tamped, no preferred orientation should occur and

TABLE I. Summary of results of determinations of the a dimension of crystallites and corresponding diamagnetic mass susceptibilities.

Sample	Heat treatment (°C)	Particle size (A)	a dimension (A)	Mass susceptibility (g^{-1})
Carbolac	3000	50[a]	90	-5.1×10^{-6}
Voltex	3000	100–200[b]	100	−5.2
Kosmobile 77	3000	300–330[b]	115	−6.7
Statex A	3000	360[b]	130	−7.2
Acetylene black	3000	430[b]	110	−6.7
Lampblack	3000	1000–2000[b]	150	−7.8
P-33	3000	1500–2000[b]	350	−6.8
Thermax	3000	2500–5000[b]	500	−7.4
Thermax	2600	2500–5000[b]	340	−7.0
Thermax	2200	2500–5000[b]	180	−7.1
Thermax	1900	2500–5000[b]	90	−4.0
Thermax	1800	2500–5000[b]	80	−1.7
Thermax	1400	2500–5000[b]	40	−1.9
Petroleum coke	3000	10 000	500	−7.8
N.C. carbon rod	1200	···	50	−1.8
N.C. carbon rod	1450	···	70	−1.9
N.C. carbon rod	1500	···	72[c]	−2.5
N.C. carbon rod	1600	···	76[c]	−2.4
N.C. carbon rod	1700	···	83[c]	−3.7
N.C. carbon rod	1800	···	90	−5.4
N.C. carbon rod	1850	···	100	−6.0
N.C. carbon rod	2200	···	180	−7.4
N.C. carbon rod	2450	···	440	−7.4
N.C. graphite rod	···	···	500	−7.8

[a] R. P. Rossman and W. R. Smith, Ind. Eng. Chem. 35, 972 (1943).
[b] *Carbon Black Manual* (Witco Chemical Company, New York, 1945).
[c] Interpolated from curve of a dimension *vs* heat treatment.

[13] E. A. Kmetko, Phys. Rev. 86, 651 (1952).

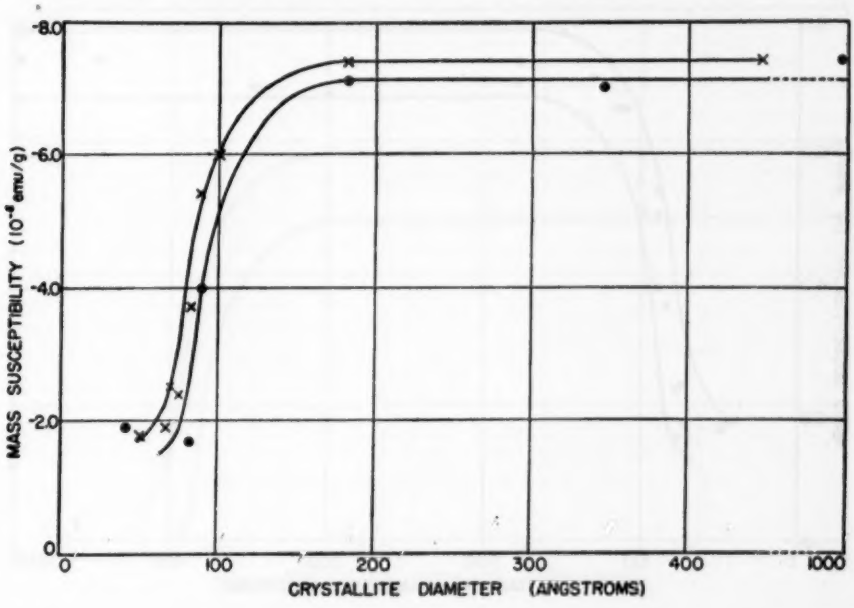

FIG. 4. Mass susceptibilities at room temperature of National Carbon rods (×) and Thermax carbon black powder (o) heat treated to different temperatures.

no correction for orientation is necessary to the values of Table I.

The value of the *a* dimension for Carbolac found in this work is larger than the average particle size as specified by the manufacturer. The 002 line of the x-ray pattern of graphitized Carbolac consists of a broad line with a sharp peak superimposed on it showing that it is a mixture of two very different size distributions. These large crystallites give a relatively larger contribution to the x-ray scattering since the contribution to the x-ray scattering is proportional to their mass. The average particle size is generally determined from average diameter so that the x-ray size measurement weights the larger sizes more heavily. In addition, Carbolac is manufactured by an oxidation of larger particles so there is a greater probability for graphitized Carbolac particles to consist of a single crystallite than for other carbon blacks where the average crystallite size is about $\frac{1}{3}$ to $\frac{1}{4}$ of the particle size.

In Fig. 4, the susceptibilities at room temperature of the series of National Carbon rods and the series of Thermax carbon blacks heated to different temperatures, are plotted as a function of the crystallite diameter. The two curves have similar shapes but reach slightly different values at larger crystallite sizes where the susceptibility is almost constant with increasing size. In the process of extrusion of carbon rods, the coke particles become aligned and there is consequently a preferential orientation of the crystallites relative to the axis of the rod. The preferred orientation will be with the crystallite planes parallel to the axis of the rod but randomly oriented about this axis. The partial alignment occurring in extrusion might explain why slightly higher values are obtained for commercial graphite rods than for carbon blacks and polycrystalline graphites with no preferential alignment.

There is also a small difference in the size range at which the rapid increase of susceptibility occurs. This might be due to the two series of samples having a somewhat different distribution of sizes, so that the averaging in the x-ray size determination would give the same average crystallite diameter, but the averaging in the susceptibility because of its different statistical weighting of the sizes would give a little different average value for susceptibility. The very similar behavior of these two series of samples suggests, however, that the susceptibility is primarily a function of crystallite size with possibly small differences arising from variations in structure due to the origin of the carbons. In Fig. 5, the susceptibilities of all the carbons investigated regardless of origin are plotted against crystallite diameter and they all fall in a narrow band which lends further support to the view that the susceptibility is in some unknown way depending on crystallite size.

DISCUSSION

As can be seen from Fig. 5, where the mass susceptibilities at room temperature of all the carbons investigated are plotted as a function of crystallite diameter, almost all the change in susceptibility from the value for very small crystallites to that for very large crystals occurs in the relatively small range of crystallite diameters from 50A–150A. This is in disagreement with the work of Rao[2] and of Gregory[3] who found that susceptibility of large crystals of natural graphite already starts decreasing when they are reduced to a size of one micron by grinding. They also found a decrease in the susceptibility of Bi ground to a similar size. It was

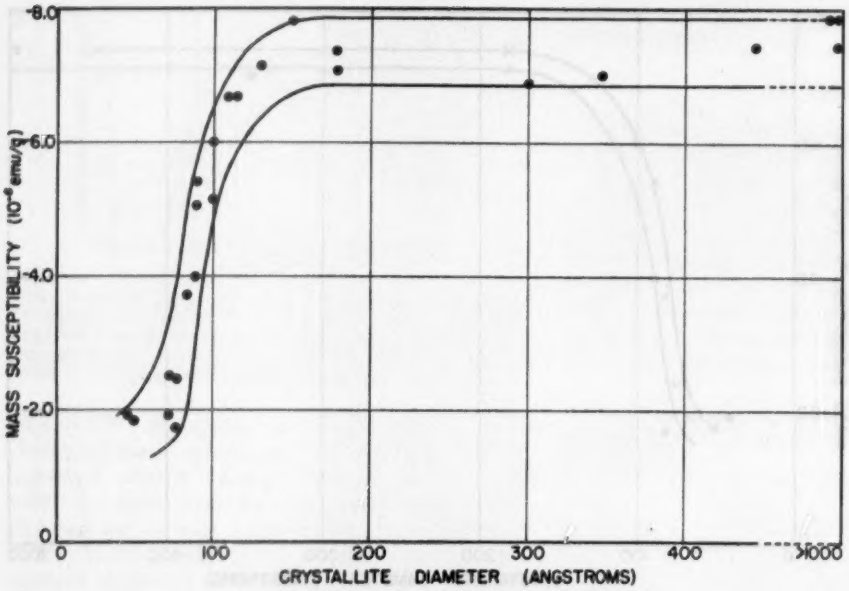

FIG. 5. Mass susceptibilities at room temperature of all carbons investigated.

shown later that the decrease of susceptibility of Bi was due to surface films of oxides and other impurities introduced by the grinding process.[14] Since it is known that oxide surface films are formed on carbon when it is ground in air,[15] it seems likely that such surface films were responsible for the decrease of susceptibility of small crystals of graphite in their experiments. The dependence of susceptibility on size found in the present investigation also disagrees with Miwa[4] who found a linear increase of susceptibility with size up to 70A (c dimension). His results, however, can only be considered as qualitative since he did not determine the crystallite size for the carbons whose susceptibility he investigated, but used the values quoted by other investigators for carbons of the same type but probably of different origin.

The recent work of Wynne-Jones, Blayden, and Iley,[7] on the other hand, is in general agreement with the results reported here in the size range from 50A–100A, with several noteworthy differences however. They found that for carbons produced by coking cellulose, dibenzathrone, and a pitch extract, the susceptibility increases slowly with size from -0.5×10^{-6}/g at 20A up to about -1.0×10^{-6}/g at 40A, then more and more rapidly for diameters in the range from 40A–100A. The susceptibilities of these three particular carbons are somewhat lower at small sizes than the carbons investigated in the present work. The susceptibility at such small sizes, however, depends strongly on the percentage of the carbon which is in the form of condensed molecules. In the coking of the particular organic materials which they used, the aromatization process is probably less rapid due to the formation of cross links between rings which hamper the growth of larger condensed ring molecules. Furthermore, the region of crystallite sizes where they found the fast increase of susceptibility is shifted towards somewhat small sizes relative to those found in this work. They did not observe the leveling off of the susceptibility at values about 150A since their crystallite size range extended only to 100A. Their curves for the three types of carbons do not overlap but are shifted somewhat relative to one another along the horizontal axis and this led them to conclude that there is no general relationship between susceptibility and crystallite size. Such small differences between carbons with different structures are certainly to be expected and a study of them might present an interesting field for further investigation. However, the existence of such a general dependence on size as given in Fig. 5 is very significant and requires an explanation.

The anisotropic component of susceptibility of condensed aromatic systems seems to be well understood at present. Pauling[16] was the first to propose an explanation of this large anisotropy. He assumed that in setting up the magnetic field, circulating currents are induced around the aromatic rings in addition to the regular induced currents within atoms responsible for ordinary atomic diamagnetism. The induced currents represent something like a Larmor precession of the π electrons and the calculated magnetic moment increases very rapidly with size. This semiclassical model gives a value more than 50 percent too large in the case of four condensed rings (pyrene) and a value equal to that of graphite already for a molecular diameter of about 40A.

[14] Bhatnagar, Varma, and Anwar Al-Hag, Kolloid-Z. 78, 9 (1937).

[15] R. H. Savage and C. Brown. J. Am. Chem. Soc. 70, 2362 (1948).

[16] L. Pauling, J. Chem. Phys. 4, 673 (1935).

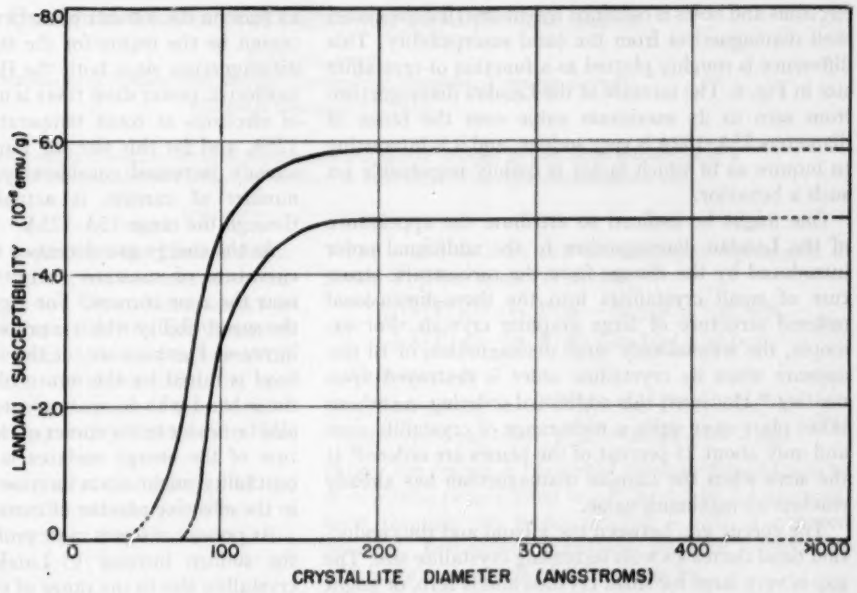

FIG. 6. Landau diamagnetism obtained by subtracting closed-shell diamagnetism from susceptibility curve, Fig. 5.

London[17] and Brooks[18] made quantum-mechanical calculations for aromatic systems and London, for example, obtained values in good agreement with experiment for systems containing up to four rings. These calculations show that the anisotropic component of susceptibility is very nearly proportional to the number of benzene rings with small deviations due to differences in arrangement of the rings. Experimentally, the proportionality to the number of benzene rings holds up to a ten-ring system (ovalence[19]). It has been pointed out[6] that this is just what one would expect since the resonance electrons in these aromatic molecules just fill a Brillouin zone and for a full zone the diamagnetism should be proportional to the molecular volume or in this case to the number of benzene rings. Assuming that this result can be extrapolated to larger systems, it is found that the average susceptibility of the π electrons plus the atomic susceptibility for randomly arranged planes of 15A diameter is -1.6×10^{-6}/g, for planes 25A in diameter, -1.8×10^{-6}/g, and for the limiting case of very large crystals (two carbon atoms per ring), -2.0×10^{-6}/g. The planes in crystallites of carbon are not all of the same size and shape. Furthermore, some of the carbon is in the form of a disorganized phase dispersed in the intercrystalline regions. Therefore, these extrapolated values will represent an upper limit for the total closed shell diamagnetism at a given size, the lower limit being -0.5×10^{-6}/g, the value for completely disorganized carbon atoms. The relative amount of carbon in the disorganized form will undoubtedly depend on the starting material so that for crystallite sizes up to 50A, the susceptibility will fall somewhere between these two limits; in fact, this is what is always observed in experiments.

It is evident that the closed-shell diamagnetism can account for only about a quarter of the total susceptibility of randomly arranged large crystals of graphite. Therefore, the remaining three-quarters must have a different origin. Ganguli and Krishnan[1] assumed that the anomalous susceptibility of graphite is due to a Landau-Peierls diamagnetism of the π electrons which form a two-dimensional gas with a low temperature of degeneracy. In order to fit the theoretical curve with the experimental data, they further assumed that electrons have effective masses equal to the electronic mass for directions of motion in the plane, but a very large effective mass in the direction perpendicular to the plane. However, their theory raised some serious objections in regard to specific heat and paramagnetism. A modification of the Ganguli-Krishnan theory which considers the Landau diamagnetism of holes and excited electrons removes these difficulties and gives a reasonable agreement with the experimental data.[6]

As mentioned above, for sizes below 50A the measured susceptibilities are always below the limiting value for the closed-shell diamagnetism of aromatic molecules and it seems probable, therefore, that in this region the Landau diamagnetism is zero or at least very small. For sizes of 75A and above, the disorganized carbon phase is not present, having been used up in the growth of turbostratic crystallites. This is shown by the disappearance of the strong diffuse scattering occurring in the x-ray patterns of all baked carbons which has been observed by Franklin[9] and in our laboratory. Therefore, above 75A the Landau diamagnetism of excited

[17] F. London, J. phys. radium **B**, 397 (1937).
[18] H. Brooks, J. Chem. Phys. **8**, 939 (1940) and **9**, 463 (1941).
[19] Akamatu, Inokuchi, and Handa, Nature **168**, 520 (1951).

electrons and holes is obtained by subtracting the closed shell diamagnetism from the total susceptibility. This difference is roughly plotted as a function of crystallite size in Fig. 6. The increase of the Landau diamagnetism from zero to its maximum value over the range of diameters 75A–150A is very sudden, and it is interesting to inquire as to which factor is mainly responsible for such a behavior.

One might be inclined to attribute the appearance of the Landau diamagnetism to the additional order introduced by the change from the turbostratic structure of small crystallites into the three-dimensional ordered structure of large graphite crystals. For example, the anomalously large diamagnetism of Bi disappears when its crystalline order is destroyed upon melting.[20] However, this additional ordering in carbons takes place over quite a wide range of crystallite sizes and only about 25 percent of the planes are ordered[9] at the sizes when the Landau diamagnetism has already reached its maximum value.

The energy gap between the π band and the conduction band decreases with increasing crystallite size. The gap is very large for small crystals and is zero, or might even be negative (slight overlap of bands)[21] for large crystals of graphite. One might ask if the steep rise in Landau diamagnetism is not directly connected with thermal activation of electrons into the conduction band. In order to examine this explanation, the behavior of the Hall effect[22] and thermoelectric power[23] as a function of crystallite size has to be considered. For carbons with crystallite size of 50A the Hall coefficient and thermoelectric power are both negative, since so many π electrons are trapped in surface states that the Fermi level is depressed below the inflection curve on the energy surface. As the crystallite size is increased the number of surface traps becomes smaller, and the electrons released from these traps return to the π band thus raising the Fermi level. At crystallite sizes of approximately 75A, the Fermi surface is located close to the inflection curve and in this range the Hall coefficient and the thermoelectric power change sign. For larger sizes the filling of the π band increases steadily. As a result, the effective number of carriers decreases, and the Hall coefficient and thermoelectric power increase up to a positive maximum at sizes of about 125A. For sizes larger than 125A, the Hall coefficient and thermoelectric power start decreasing since the activation energy becomes progressively smaller and consequently more and more electrons are thermally excited into the conduction band. It is evident that the increase in the number of carriers by thermal excitation cannot be the reason for the steep increase in Landau diamagnetism since both the Hall coefficient and thermoelectric power show there is no appreciable excitation of electrons at room temperature for sizes less than 125A, and for this size the Landau diamagnetism has already increased considerably. In fact, the effective number of carriers is actually decreasing slightly through the range 75A–125A.[6]

As the energy gap decreases with increasing size, the curvature of constant energy curves might increase near the zone corners.[6] For a given filling of the zone, the susceptibility which is proportional to $\alpha_1\alpha_2$ will thus increase. Furthermore, at the same time as the Fermi level is raised by the return of trapped electrons into the π band with increasing size, the Fermi surface will also be nearer to the corner of the zone where the curvature of the energy surfaces is greater; thus the susceptibility might again increase in spite of the decrease in the effective number of carriers.

At present it seems most probable to the author that the sudden increase in Landau diamagnetism with crystallite size in the range of sizes 75A–150A is due to the combined influence of these two factors, that is of the increase of curvature of the energy surfaces near the zone corners and the raising of the Fermi level into regions of higher curvature. An investigation of the susceptibilities of residue compounds of carbons with bisulfate ions and of the temperature dependence of susceptibility of carbons with different crystallite sizes is currently being performed and it is hoped that it will furnish some information as to the relative importance of these two factors.

It seems very significant, however, that the susceptibility reaches almost its maximum value at 150A and stays approximately constant from there on over such a wide range of crystallite sizes, a range in which the Hall effect, thermoelectric effect, and electric resistivity vary so strongly. Of course, due to a lack of knowledge of the exact shape of the energy surfaces, the whole discussion can be only qualitative and the explanation of all details is at present impossible. On the other hand, it appears possible that the whole approach to the theoretical explanation will have to be strongly modified in the light of the results of the recent work on Landau-Peierls diamagnetism as reported by Dingle,[24] Adams,[25] and others.

The author wishes to express his thanks to Professor S. Mrozowski, who suggested the problem, for his continued advice and encouragement. The x-ray spectrometer was obtained with the assistance of a grant from the Research Corporation.

[20] N. F. Mott and H. Jones, *Theory of the Properties of Metals and Alloys* (Clarendon Press, Oxford, 1936).
[21] S. Mrozowski, J. Chem. Phys. 21, 492 (1953).
[22] E. A. Kmetko, J. Chem. Phys. 21, 2152 (1953).
[23] E. E. Loebner, Phys. Rev. 84, 153 (1951); Phys. Rev. 86, 1056 (1952).
[24] R. B. Dingle, Proc. Roy. Soc. (London) 216, 118 (1953).
[25] E. N. Adams II, Phys. Rev. 89, 633 (1952).

A Quantum-Mechanical Definition of Refractive Index

S. M. NEAMTAN
Department of Mathematical Physics, University of Manitoba, Winnipeg, Canada
(Received December 21, 1953)

The refractive index is defined in terms of momentum and energy changes in first-order transitions of a system consisting of the particles of a refractive medium in interaction with the electromagnetic field.

IN a previous article,[1] a quantum-mechanical definition was given for the dielectric constant of a nonmagnetic material, and it was shown that subject to certain restrictive conditions it led to the Kramers-Heisenberg dispersion formula. It is the purpose of the present paper to show that the above definition, considered now as a definition of the square of the refractive index, is of much greater generality than was at first apparent.

As in I, we consider a system S consisting of the electromagnetic field and the atomic particles of a refractive medium in interaction. The Hamiltonian for this system has the form

$$H=\sum_\lambda 2\nu_\lambda q_\lambda^* q_\lambda+\sum_i f_i\left(\mathbf{p}_i-\frac{e_i}{c}\mathbf{A}(\mathbf{r}_i)\right)+\sum_{i\neq j}\frac{e_ie_j}{|\mathbf{r}_i-\mathbf{r}_j|}, \quad (1)$$

where the form of the function f_i will depend on the nature of the ith particle and on whether it is treated relativistically or nonrelativistically. In particular, f_i may be a function of spin variables but is to depend on positional variables only through the term $\mathbf{A}(\mathbf{r}_i)$ in its agrument. The momentum of the system S is given by

$$\mathbf{P}=\sum_\lambda 2\nu_\lambda\boldsymbol{\kappa}_\lambda q_\lambda^* q_\lambda+\sum_i \mathbf{p}_i. \quad (2)$$

To verify that this is the correct expression, it is to be noted first that it has the correct form for the limiting case of no interaction between particles and radiation. It then remains only to show that $\mathbf{P}$ is a constant of the motion, that is, that $\mathbf{P}$ commutes with H. With the aid of the relations,

$$\nabla\cdot\mathbf{A}=0, \quad (3)$$

$$[q_\lambda,q_\mu^*]=(\hbar/2\nu_\lambda)\delta_{\lambda\mu}, \quad (4)$$

and

$$[q_\lambda,q_\mu]=[q_\lambda^*,q_\mu^*]=0, \quad (5)$$

it is easily verified that

$$[\mathbf{P},\mathbf{p}_i-(e_i/c)\mathbf{A}(\mathbf{r}_i)]=0; \quad (6)$$

then since we also have

$$\left[\sum_i \mathbf{p}_i,\ \sum_{i\neq j}\frac{e_ie_j}{|\mathbf{r}_i-\mathbf{r}_j|}\right]=0, \quad (7)$$

it follows that

$$[\mathbf{P},H]=0. \quad (8)$$

With the aid of (2), (4), and (5), we also obtain the relations:

$$[\mathbf{P},q_\lambda^*]=\hbar\boldsymbol{\kappa}_\lambda q_\lambda^* \quad \text{and} \quad [\mathbf{P},q_\lambda]=-\hbar\boldsymbol{\kappa}_\lambda q_\lambda. \quad (9)$$

[1] S. M. Neamtan, Phys. Rev. **92**, 1362 (1953). The notation used will be that of the above article, to be referred to henceforth as I.

Let n and m label two simultaneous eigenstates of H and $\mathbf{P}$. Then from (9) we find

$$\begin{aligned}\{\langle n|\mathbf{P}|n\rangle-\langle m|\mathbf{P}|m\rangle-\hbar\boldsymbol{\kappa}_\lambda\}\langle n|q_\lambda^*|m\rangle=0,\\ \{\langle n|\mathbf{P}|n\rangle-\langle m|\mathbf{P}|m\rangle+\hbar\boldsymbol{\kappa}_\lambda\}\langle n|q_\lambda|m\rangle=0.\end{aligned} \quad (10)$$

Thus we have

$$\begin{aligned}\langle n|q_\lambda^*|m\rangle=0 \text{ unless } \langle n|\mathbf{P}|n\rangle-\langle m|\mathbf{P}|m\rangle=\hbar\boldsymbol{\kappa}_\lambda,\\ \langle n|q_\lambda|m\rangle=0 \text{ unless } \langle n|\mathbf{P}|n\rangle-\langle m|\mathbf{P}|m\rangle=-\hbar\boldsymbol{\kappa}_\lambda.\end{aligned} \quad (11)$$

If now the system is perturbed through the addition to H of a term proportional to q_λ^* (as would, for example, arise through the interaction of S with a charged particle), first-order transitions may be induced in which the momentum of S increases by

$$\Delta\mathbf{P}_\lambda=\hbar\boldsymbol{\kappa}_\lambda, \quad (12)$$

while transitions induced by a term proportional to q_λ will correspond to a decrease in momentum of the same amount. In the absence of a refractive medium, such transitions correspond to the emission (or absorption) of a photon of momentum $\hbar\boldsymbol{\kappa}_\lambda$. For the system S, therefore, we must interpret such transitions as corresponding to the emission (or absorption) in the medium of radiation with the same propagation vector $\boldsymbol{\kappa}_\lambda$.

Let α be any dynamical variable of the system S. Then in a representation in which H is diagonal, the matrix elements of α and $\ddot{\alpha}$ satisfy the relation:

$$\langle n|\ddot{\alpha}|m\rangle+\omega_{nm}^2\langle n|\alpha|m\rangle=0, \quad (13)$$

where

$$\hbar\omega_{nm}=\langle n|H|n\rangle-\langle m|H|m\rangle. \quad (14)$$

In particular, taking $\alpha=q_\lambda^*$, we have

$$\langle n|\ddot{q}_\lambda^*|m\rangle+\omega_{nm}^2\langle n|q_\lambda^*|m\rangle=0. \quad (15)$$

In a vacuum, the wave equation for the vector potential $\mathbf{A}$, namely

$$\nabla^2\mathbf{A}-(1/c^2)\partial^2\mathbf{A}/\partial t^2=0, \quad (16)$$

yields

$$\langle n|\ddot{q}_\lambda^*|m\rangle+c^2\kappa_\lambda^2\langle n|q_\lambda^*|m\rangle=0. \quad (17)$$

If now (15) is to correspond to a wave equation in a refractive medium, we must write

$$c^2\kappa_\lambda^2/\omega_{nm}^2=\mu^2, \quad (18)$$

where μ is the refractive index.

We have thus been led to the following definition. If, in a first-order transition of the system S corresponding to the emission or absorption of radiation with propagation vector $\boldsymbol{\kappa}_\lambda$, the change in momentum is $\Delta\mathbf{P}$ and the change in energy is ΔW, the refractive index is given by

$$\mu=c|\Delta\mathbf{P}|/|\Delta W|. \quad (19)$$

Studies on Localized Orbitals*

HARVEY WINSTON
Hughes Research and Development Laboratories, Culver City, California
(Received December 28, 1953)

Wannier functions for free electrons in one dimension are discussed in some detail. The effects of the choice of lattice points and unit cell size are discussed, and details of the symmetry properties of the one-dimensional nearly-free-electron wave functions are given. The general relation between symmetry in the local site and three-dimensional space group symmetry is given. The possibilities for localized orbital treatments of metals and perturbed lattices are considered, and the electronic structure of CdSb is examined in terms of localized orbitals. The term *site orbital* is proposed to include the special cases of Wannier functions and Lennard-Jones equivalent orbitals.

INTRODUCTION

A NUMBER of different kinds of localized orbitals have been discussed in recent years. Wannier functions have been given a central position in Slater's[1] work; they are defined by

$$a(\mathbf{q}-\mathbf{Q}_n)=N^{-\frac{1}{2}}\sum_{\mathbf{p}_0} u_{\mathbf{p}_0}(\mathbf{q})\exp(-i\mathbf{p}_0\cdot\mathbf{Q}_n/\hbar), \quad (1)$$

in which $\mathbf{p}_0$ runs over *all* the crystal momenta of a band, not just those for which the corresponding functions $u_{\mathbf{p}_0}$ are occupied. There are some ambiguities in this definition for multidimensional cases, with respect particularly to the question of which functions are to be included in a band.

Lennard-Jones' school[2] has placed special emphasis on the equivalent orbital (EO). The EO work has dealt mostly with molecules, but there has been an application[3] to crystals of the diamond structure. EO are defined as a set of orbitals spanning the same function space as some set of filled molecular orbitals or crystal orbitals, but having the property that their constant value surfaces are of identical shape, differing only in location and/or orientation. An antisymmetrized determinant of EO can be shown to be the same function of electron coordinates and spins as the antisymmetrized determinant of the molecular or crystal orbitals from which the EO are formed. As Lennard-Jones has shown, this property arises from the degrees of freedom possessed by the solutions of a multielectron Schrödinger equation according to the Hartree-Fock method.

It proves convenient to make use of a kind of orbital for which we shall introduce the designation "site orbital" (SO). These SO are like EO except that they need not be limited to being linear combinations of filled orbitals; EO are thus special cases of SO. SO in general do not have the EO property mentioned at the end of the last paragraph. The number of SO in a set is the product of the number of equivalent sites of a kind in the molecule or crystal by the degeneracy of the site orbital in the site symmetry.

The great advantage of the EO approach is that certain features of the EO theory correspond to the chemical ideas of bond, resonance, and atomic orbitals left unaffected by bonding. Discussing molecular orbital theory in EO terms enables contact to be made with these chemical ideas, thus clarifying the connection between theory and experiment. Application of the same kind of idea in crystals has not progressed as far.

WANNIER FUNCTIONS FOR FREE ELECTRONS

It seems worth while to examine the Wannier functions for free electrons in more detail than previous authors as a prelimary to the discussion of symmetry properties of localized orbitals. We begin by setting up Wannier functions for the bands in a one-dimensional, almost-free-electron situation. Our one-dimensional crystal has the unit cell length α, and is N unit cells long, with a total length L given by $N\alpha$. Except for functions with energies in the immediate neighborhood of the band gap, the eigenfunctions are

$$L^{-\frac{1}{2}}\exp(ip\cdot q/\hbar). \quad (2)$$

As the perturbation is decreased and the potential becomes more nearly that for completely free electrons, the neighborhoods of the band gaps are contracted, the gaps themselves decrease, and the functions at the band gaps become, for the nth band, as is well known,

$$\left(\frac{2}{L}\right)^{\frac{1}{2}}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}\left\{\begin{matrix}n\\ n-1\end{matrix}\right\}\frac{\pi}{\alpha}q. \quad (3)$$

In a perfectly free-electron case, there are no band gaps, and any length at all might be chosen for α. The assignment of functions to bands depends on the choice of α; we will take advantage of this when we discuss the effect of choosing unit cell lengths which are multiples of the primitive length.

The Wannier function for the lth band is given by

$$a^l(q-Q_n)=N^{-\frac{1}{2}}\sum_{p_0} u_{p_0}{}^l(q)\exp(-ip_0\cdot Q_n/\hbar). \quad (4)$$

* Supported in part by the Signal Corps.
[1] J. C. Slater, Phys. Rev. 76, 1592 (1949); 87, 807 (1952), and references therein.
[2] J. E. Lennard-Jones, Proc. Roy. Soc. (London) A198, 1 (1949) and many later papers.
[3] G. G. Hall, Phil. Mag. 43, 338 (1952).

Evidently the form of $a(q-Q_n)$ depends on the choice of phases for the eigenfunctions. The choice made in expressions (2) is only one of many possibilities; in a later section we consider the implications of this, meanwhile keeping to the phases of Eq. (2). Substitution in Eq. (4) gives on this basis

$$a^l(q-Q_n)=(NL)^{-\frac{1}{2}}\sum_{p=(l-1)h/2a}^{lh/2a}\{\exp[ip(q-Q_n)/h] +\exp[-ip(q-Q_n)/h]\}. \quad (5)$$

The p's in the summation are to be chosen by the periodic boundary condition

$$p=(m/N)(h/a), \quad (6)$$

with m integral. It is to be noted that Eq. (5) is really inconsistent with Eq. (4) in terms of order $N^{-\frac{1}{2}}$; the band edge functions are really given by Eq. (3), and the band edge functions used in (4) are

$$L^{-\frac{1}{2}}\cos\left\{\begin{matrix}l-1\\ l\end{matrix}\right\}\frac{\pi}{a}q. \quad (7)$$

We may expect some difficulties in this connection.

In preparation for replacing the summation in Eq. (5) by integration, we rewrite:

$$a^l(q-Q_n)=(NL)^{-\frac{1}{2}}\{\exp[\pi i(l-1)(q-Q_n)/a] \times\sum_{m=0}^{N/2}\exp[2\pi im(q-Q_n)/aN] +\exp[-\pi i(l-1)(q-Q_n)/a] \times\sum_{m=0}^{N/2}\exp[-2\pi im(q-Q_n)/aN]\}. \quad (8)$$

We now replace summation over m by integration over dm, which is equivalent to going to infinite N and keeping the summation. The result is:

$$(NL)^{\frac{1}{2}}[2\pi i(q-Q_n)/Na]a^l(q-Q_n) =\{\exp[\pi i(l-1)(q-Q_n)/a]\} \times\{-1+\exp[\pi i(q-Q_n)/a]\} -\{\exp[-\pi i(l-1)(q-Q_n)/a]\} \times\{-1+\exp[-\pi i(q-Q_n)/a]\}, \quad (9)$$

$$a^l(q-Q_n)=a^{-\frac{1}{2}}[a/\pi(q-Q_n)]\{\sin[l\pi(q-Q_n)/a] -\sin[(l-1)\pi(q-Q_n)/a]\} \quad (9A)$$

$$=[2a^{\frac{1}{2}}/\pi(q-Q_n)]\sin[\pi(q-Q_n)/2a] \times\cos[\pi(l-\tfrac{1}{2})(q-Q_n)/a]. \quad (9B)$$

All this is a generalization of Wannier's original work[4] and agrees with it, of course, in the case of $l=1$ worked out by him. Wannier functions should have the property of orthonormality in the range $-\infty\leftrightarrow\infty$ for the indices l and n; this is shown to be the case in an appendix.

[4] G. H. Wannier, Phys. Rev. 52, 191 (1937).

It is instructive to compare the Wannier functions for the lth band with the eigenfunctions of an electron in a box centered at Q_n with total length a. The lth such function is

$$\psi_l=\sqrt{2}a^{-\frac{1}{2}}\cos[\pi l(q-Q_n)/a]; \quad E_l=h^2l^2/8ma^2. \quad (10)$$

Equation (9B) shows that the lth Wannier function is similar to Eq. (10); it has $(l-\frac{1}{2})$ in place of l and is modulated by $[(2a)^{\frac{1}{2}}/\pi(q-Q_n)]\sin[\pi(q-Q_n/2a]$. The energies of the lth band run from $h^2(l-1)^2/8ma^2$ to $h^2l^2/8ma^2$.

There is another choice of phases for the summation in Eq. (5) which leads to particularly simple results. In this choice, functions with $p_0>0$ are positive real at $q=0$; functions with $p_0<0$ are negative real at the origin. In this case Eq. (8) will have the plus sign connecting the summations in curly brackets replaced by minus sign, and Eqs. (9A) and (9B) become

$$a_-{}^l(q-Q_n)=[a^{\frac{1}{2}}/\pi i(q-Q_n)]\{\cos[\pi l(q-Q_n)/a] -\cos[\pi(l-1)(q-Q_n)/a]\}, \quad (9A')$$

$$a_-{}^l(q-Q_n)=[2ia^{\frac{1}{2}}/\pi(q-Q_n)]\sin[\pi(q-Q_n)/a] \times\sin[(l-\tfrac{1}{2})\pi(q-Q_n)/a]. \quad (9B')$$

We shall henceforth refer to the a's of Eqs. (9) as a_+ to distinguish them from a_- of Eqs. (9'). Since the sets $a_+(q-Q_n)$ and $a_-(q-Q_n)$ can be used to express the eigenfunctions of the same band, we expect to find a relation between them (probably correct up to terms of order $N^{-\frac{1}{2}}$). This relation, derived below, turns out to be valid in the general one-dimensional case and is not limited to the free-electron potential.

Let us try to express $a_-(q-Q_n)$ in terms of the set $a_+(q-Q_n)$. We see immediately that

$$a(q-Q_n)=N^{-\frac{1}{2}}\sum_{p_0}f(p_0)u_{p_0}(q)\exp(-ip_0\cdot Q_n/h), \quad (11)$$

where $f(p_0)$ is plus one for $p_0>0$ and minus one for $p_0<0$, and the eigenfunctions u_{p_0} incorporate the phase choice of the original equations (9). In general,

$$u_{p_0}(q)=N^{-\frac{1}{2}}\sum_{Q_n}a_+(q-Q_n)\exp(ip_0\cdot Q_n/h). \quad (12)$$

We substitute (12) into (11) to get

$$a_-(q-Q_n)=N^{-1}\sum_{Q_k}\sum_{p_0}f(p_0)a_+(q-Q_k) \times\exp[ip_0\cdot(Q_k-Q_n)/h]. \quad (13)$$

Replacing the summation over p_0 by integration leads finally to

$$a_-(q-Q_n)=-i\sum_{Q_k}a_+(q-Q_k) \times\{\cos[\pi(Q_k-Q_n)/a]-1\}[a/\pi(Q_k-Q_n)]. \quad (14)$$

When we substitute our a_+ into Eq. (12), the result should be $u_{p_0}(q)$; if a_- is used, we should get $u_{p_0}(q)$ for $p_0>0$, $-u_{p_0}(q)$ for $p_0<0$, or $f(p_0)u_{p_0}(q)$. Before check-

ing this, it is worth while to see what exceptions there might be. Suppose that our $a(q-Q_n)$ differs from the true Wannier function A by a term of order $N^{-\frac{1}{2}}$:

$$A^l(q-Q_n)=a^l(q-Q_n)+N^{-\frac{1}{2}}X. \tag{15}$$

X may be taken to be of the form

$$\sum_{d,p_0} c^d(p_0)u_{p_0}{}^d(q)\exp(-ip_0\cdot Q_n/\hbar). \tag{16}$$

In carrying out the operations of Eq. (12), we find that

$$u_{p_0}{}^l(q)=N^{-\frac{1}{2}}\sum_{Q_n} a^l(q-Q_m)\exp(ip_0\cdot Q_n/\hbar) + \sum_d c^d(p_0)u_{p_0}{}^d(q). \tag{17}$$

Hence, it may be that for certain p_0 (for which $c(p_0)\neq 0$),

$$N^{-\frac{1}{2}}\sum a^l(q-Q_n)\exp(ip_0\cdot Q_n/\hbar) = u_{p_0}{}^l(q)-\sum_d c^d(p_0)u_{p_0}{}^d(q). \tag{18}$$

In other words, the result of applying Eq. (12) to $a(q-Q_n)$ yields a result which is characterized by p_0, but it is not the normalized eigenfunction u_{p_0} unless all $c^d(p_0)=0$. This comes about because Eq. (12) picks up the term of order $N^{-\frac{1}{2}}$ in $a(q-Q_n)$; we have to be careful because we are in effect taking a sum of N terms of order $N^{-\frac{1}{2}}N^{-\frac{1}{2}}=N^{-1}$, and this sum need not vanish. It is plausible, however, to expect this sort of difficulty at only a finite number of points p_0 (a vanishingly small fraction of reciprocal space as $N\to\infty$), since our a's are actually orthonormal. If $c^d(p_0)$ were not a sum of δ functions around a limited number of p_0's, it would be hard to see how orthonormality could be preserved. Another way of putting this is that if $c^d(p_0)$ were different from zero over a finite range of p_0's, the term in X would no longer really be of order $N^{-\frac{1}{2}}$.

Keeping the discussion of the preceding paragraph in mind, we proceed to evaluate Eq. (12) for our a's at the points $p_0=0$, $h/2\alpha$, and $h/4\alpha$. Let $q/\alpha=x$, and $Q_n=n\alpha$. Then, for $p_0=0$,

$$N^{-\frac{1}{2}}\sum_n a_\pm{}^l(q-Q_n)=\frac{1}{(N\alpha)^{\frac{1}{2}}}\sum_n \frac{2}{\pi(x-n)} \times\sin[\tfrac{1}{2}\pi(x-n)]\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}[(l-\tfrac{1}{2})\pi(x-n)]. \tag{19}$$

Note that for even n

$$\sin[\tfrac{1}{2}\pi(x-n)]\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}[(l-\tfrac{1}{2})\pi(x-n)] =\sin(\tfrac{1}{2}\pi x)\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}[(l-\tfrac{1}{2})\pi x],$$

and for odd n it is given by

$$\cos(\tfrac{1}{2}\pi x)\begin{Bmatrix}(-)^l\sin\\ (-)^{l-1}\cos\end{Bmatrix}[(l-\tfrac{1}{2})\pi x].$$

It is therefore easy to break the summation of Eq. (19) into two parts to get

$$(N\alpha)^{-\frac{1}{2}}\tfrac{1}{2}\pi\left\{[\sum_{\text{even }n}(x-n)^{-1}]\sin(\tfrac{1}{2}\pi x)\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}[(l-\tfrac{1}{2})\pi x] +[\sum_{\text{odd }n}(x-n)^{-1}]\cos(\tfrac{1}{2}\pi x)\begin{Bmatrix}(-)^l\sin\\ (-)^{l-1}\cos\end{Bmatrix}[(l-\tfrac{1}{2})\pi x]\right\}.$$

We now make use of the relationships

$$\sum_r (x-2r)^{-1}=\tfrac{1}{2}\pi\cot(\tfrac{1}{2}\pi x),$$

$$\sum_r (x-1-2r)^{-1}=-\tfrac{1}{2}\pi\tan(\tfrac{1}{2}\pi x)$$

to evaluate the sums, and obtain[5]

$$\begin{aligned}(N\alpha)^{-\frac{1}{2}}&\left\{\cos(\tfrac{1}{2}\pi x)\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}[(l-\tfrac{1}{2})\pi x]\right.\\ &\left.-\sin\tfrac{1}{2}\pi x\begin{Bmatrix}(-)^l\sin\\ (-)^{l-1}\cos\end{Bmatrix}[(l-\tfrac{1}{2})\pi x]\right\}\\ &=(N\alpha)^{-\frac{1}{2}}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(l\pi x)\qquad\text{for } l \text{ even},\\ &=(N\alpha)^{-\frac{1}{2}}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}[(l-1)\pi x]\quad\text{for } l \text{ odd}.\end{aligned} \tag{20}$$

The case of $h/2\alpha$ works out to

$$\begin{aligned}&(N\alpha)^{-\frac{1}{2}}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}[(l-1)\pi x]\quad\text{for } l \text{ even};\\ &(N\alpha)^{-\frac{1}{2}}\begin{Bmatrix}\cos\\ \sin\end{Bmatrix}(l\pi x)\qquad\text{for } l \text{ odd}.\end{aligned} \tag{21}$$

For $p_0=h/4\alpha$, it turns out that the sums involved must be split up into four partial sums, and the final result is, for $l=1$, using either a_+ or a_-,

$$(N\alpha)^{-\frac{1}{2}}\exp(\tfrac{1}{2}\pi i x). \tag{22}$$

Equation (22) is the result we expect for almost all p_0's: it shows that Eq. (12) yields the true normalized eigenfunctions. The results summarized in Eqs. (21) and (22) are not normalized eigenfunctions; they lack a factor of $\sqrt{2}$. Moreover, it is clear that to get the result even proportional to the correct eigenfunction,

[5] The author is indebted to Mr. T. O. Woodruff for pointing out this relationship.

a_+ must be used when the band-edge functions are cosines, and a_- when they are sines. At other p_0's it is immaterial whether a_+ or a_- is employed, but at $p_0=0$ and $h/2\alpha$, the terms in $N^{-\frac{1}{2}}$ make trouble.

In this connection it is of interest to see what happens when we use Eq. (14) to express a_- in terms of a_+ and then try to evaluate $u_{p0}(q)$ for $p_0=0$:

$$N^{-\frac{1}{2}}\sum_n a_-(q-Q_n)$$
$$=-iN^{-\frac{1}{2}}\sum_n\sum_k \frac{\{\cos[\pi(Q_k-Q_n)/\alpha]\}-1}{\pi(Q_k-Q_n)/\alpha}a_+(q-Q_n)$$
$$=-iN^{-\frac{1}{2}}\sum_n\left\{\sum_k \frac{\cos[\pi(Q_k-Q_n)/\alpha]-1}{\pi(Q_k-Q_n)/\alpha}\right\}a_+(q-Q_n). \quad (23)$$

This expression vanishes because the coefficient of $a_+(q-Q_n)$ is a sum over Q_k of terms which are odd in (Q_k-Q_n); thus each positive contribution is cancelled by an equal and opposite one. Under no circumstances can a_+ be made to express a sine function. When we try to force the issue, only zero emerges.

CHANGING THE LATTICE POINTS

One rather arbitrary feature of the definition of Wannier functions is the choice of the lattice points as the set $\{Q_k\}$. Let us see, in the general one-dimensional case, what the effect is of using the set $\{Q_k+\Delta\}$ instead. Begin by writing the one-electron eigenfunctions in terms of the Wannier functions defined using Q_k as the set of lattice points, as in Eq. (12). Let

$$a^\Delta(q-Q_n-\Delta)=N^{-\frac{1}{2}}\sum_{p_0} u_{p0}(q) \times\exp[-ip_0\cdot(Q_n+\Delta)/\hbar]. \quad (24)$$

Then

$$a^\Delta(q-Q_n-\Delta)=N^{-1}\sum_k\sum_{p_0} a(q-Q_k) \times\exp[ip_0\cdot(Q_k-Q_n-\Delta)/\hbar]. \quad (25)$$

Summation, or rather integration, over p_0 leads to:

$$a^\Delta(q-Q_n-\Delta)=\sum_k \frac{\sin[\pi(Q_k-Q_n-\Delta)/\alpha]}{\pi(Q_k-Q_n-\Delta)/\alpha}a(q-Q_k). \quad (26)$$

Extending Eq. (26) to three dimensions offers no difficulty.

The relation (26) illustrates the lack of uniqueness of the Wannier function definition. Another way to describe the situation is to regard Eq. (26) as the formula giving the new Wannier functions appropriate to a new choice of phases in which the phase of each u_{p0} is changed by $p_0\Delta/\hbar$. The right-hand side of Eq. (26) appears highly delocalized, with contributions from Wannier functions centered at many points; but of course it is equal to the left-hand side, which contains just one Wannier function of the type a^Δ. A direct check of the last equation for our nearly-free-electron case is given in an appendix. Another possibility for the free-electron case is a phase change in each function given by $p\Delta/\hbar$ rather than by $p_0\Delta/\hbar$. The interesting relationships to which this choice leads are also dealt with in an appendix.

CHANGING UNIT CELL SIZE

Slater[6] has worked out the effect upon the Wannier functions of doubling the size of the unit length in a one-dimensional system. We shall cover approximately the same ground in this section; the freedom in manipulating Wannier functions given us by Eq. (26) makes it profitable to do so.

We have to begin with a system periodic with the period α. The Wannier function of the band which concerns us is $a(q-Q_n)$. The N different functions of q realizing by choosing $Q_n=n\alpha$ form a complete set for describing any function belonging to the band. If we now decide to double the unit cell, we will have two bands of functions spanning the same function space as our one original band. There will be $\frac{1}{2}N$ functions in each of these bands, so we see that what we are doing is just choosing another set of N linear combinations of the N original band functions. The Wannier functions for the new (half-) bands will be designated as $a^{\rm I}(q-Q_n)$ and $a^{\rm II}(q-Q_n)$, with only even values of n.

In the following equations, p_0 is a crystal momentum for the original description, and μ_0 is a crystal momentum in the double unit cell system:

$$a(q-Q_n)=N^{-\frac{1}{2}}\sum_{p_0=-h/2\alpha}^{h/2\alpha} u_{p0}(q)\exp(-ip_0\cdot Q_n/\hbar), \quad (27)$$

$$a^{\rm I}(q-Q_n)=(\tfrac{1}{2}N)^{-\frac{1}{2}}\sum_{\mu_0=-h/4\alpha}^{h/4\alpha} u_{\mu_0}'\exp(-i\mu_0\cdot Q_n/\hbar), \quad (28)$$

$$a^{\rm II}(q-Q_n)=(\tfrac{1}{2}N)^{-\frac{1}{2}}\sum_{\mu_0=-h/4\alpha}^{h/4\alpha} u_{\mu_0}''(q) \times\exp(-i\mu_0\cdot Q_n/\hbar). \quad (29)$$

Equation (27) is valid for all integral n; Eqs. (28) and (29) hold for even n only. We take $u_{\mu_0}'=u_{p_0=\mu_0}$ and $u_{\mu_0}''=u_{p_0=\mu_0\pm h/2\alpha}$ using the plus sign for $\mu_0>0$ and the minus sign for $\mu_0<0$. We note that the periodic boundary condition (P.B.C.) gives $[m/(N/2)]/(h/2\alpha)=mh/N\alpha$ as the allowed values of μ_0; hence, μ_0 assumes the same values as p_0 below $|\mu_0|=h/4\alpha$. Consideration of the foregoing relations shows fairly directly that

$$a(q-Q_n)=2^{-\frac{1}{2}}[a^{\rm I}(q-Q_n)+a^{\rm II}(q-Q_n)], \quad n \text{ even}; \quad (30)$$

$$a(q-Q_n-\alpha)=2^{-\frac{1}{2}}[a^{\rm I}(q-Q_n-\alpha)-a^{\rm II}(q-Q_n-\alpha)], \quad n \text{ even}. \quad (31)$$

[6] J. C. Slater, Phys. Rev. **87**, 807 (1952).

Note especially the sign of a^{II} in Eq. (31), which is what takes account of the difference between p_0 and μ_0. The expression $a^I(q-Q_n-\alpha)$ is to be interpreted as the function obtained from Eq. (28) by replacing Q_n with $Q_n+\alpha$. Obviously, $a(q-Q_n-\alpha)$ is equal to $a(q-Q_{n+1})$, which could be shown formally from Eq. (26). We can also use Eq. (26), modified for the double periodicity, to express the right-hand side of Eq. (31) in terms of the set with even n:

$$a(q-Q_{n+1})=2^{-1}\sum_{\text{even } k}[a^I(q-Q_k)-a^{II}(q-Q_k)]$$
$$\times\{\exp[\tfrac{1}{2}\pi i(k-n)]\}[2/\pi(k-n-1)]. \quad (32)$$

Equations (30) and (32) show us how to express the N Wannier function of the full band in terms of the 2 times $\frac{1}{2}N$ Wannier functions for the two half-bands, defined for even n only.

The converse problem, that of expressing a^I and a^{II} in terms of a, can be handled by expanding the former in series of the latter, using Eqs. (30) and (32) to evaluate the coefficients:

$$a^{I,II}(q-Q_n)=2^{-1}\sum_{\text{even } r}\Big[\delta_{rn}a(q-Q_r)$$
$$\pm a(q-Q_{r+1})\frac{\exp[\frac{1}{2}\pi i(n-r)]}{\frac{1}{2}\pi(n-r-1)}\Big]. \quad (33)$$

We use the plus sign in expanding a^I and the minus sign for a^{II}. We see very clearly that the reason for the appearance of Wannier functions from many sites in the expansion is our exclusion of half-band functions at other than even lattice points and our subsequent use of Eq. (26) to transform these away.

SYMMETRY PROPERTIES OF WANNIER FUNCTIONS

Our aim in this section is to see what can be said about Wannier functions on the basis of crystal symmetry. Previous work in this direction has been done by Slater,[6] Koster,[7] and Feuer.[8]

The definition of Wannier functions shows that they are linear combinations of eigenfunctions. Since the eigenfunctions belong to irreducible representations of the space group of the crystal, it appears that Wannier functions could be classified symmetry-wise by specifying just which irreducible representations (I.R.) of the space group characterize the eigenfunctions of the band. Another kind of symmetry characterization which might be desirable is a specification of the behavior of a Wannier function under the symmetry operations which leave invariant the point about which it is localized.

It is convenient to begin with a one-dimensional example which exhibits some of the results without complicated geometry. A case at hand is our nearly-free-electron one, which will have a space group involving translations by multiples of α along the line, and reflections in lines perpendicular to the line of the crystal at points $\frac{1}{2}n\alpha$, with n any integer. Consider the case in which the (P.B.C.) is applied to a region of the line N unit cells long. No real loss in generality is involved in assuming that N is even. Any operation of the group may be expressed in the form $\{R|\mathbf{t}\}$, where R is either the identity or a reflection in the origin, and $\mathbf{t}$ is a translation $n\alpha$. On account of the P.B.C., operations which leave the line translated by $N\alpha$ from its position at the end of a given operation are considered to be the same as the given operation. All pure translations are of the form $\{E|n\alpha\}$, and $\{E|(n+N)\alpha\}$ is the same as $\{E|n\alpha\}$. Reflections are of two types; those which leave invariant a point translationally equivalent to the origin, of the form $\{\sigma|n\alpha\}$ with n even; and those which leave invariant the points translationally equivalent to $\frac{1}{2}\alpha$, of the form $\{\sigma|n\alpha\}$ with n odd. It is to be noted that $\{\sigma|n\alpha\}$ and $\{\sigma|(n+N)\alpha\}$ are to be regarded as the same operation, since they differ only in that the line is translated by $N\alpha$ after one of these operations with respect to its position after the other. This (or these) operation(s) leaves invariant two points, at $\frac{1}{2}n\alpha$ and at $\frac{1}{2}(n+N)\alpha$.

We can get at the symmetry properties of a band by considering the character system of the reducible representation for which its functions provide a basis. These character systems are represented in Table I for some interesting cases. The four rows of entries are for, respectively, the set of even Wannier functions for a band, the set of odd Wannier functions, the set of eigenfunctions of a band for which the band-edge functions are cosines, and the set of eigenfunctions with sine-type band-edge functions. The character entries are obtained simply, as follows: $\chi(E)$ is N for all four sets since they all contain N functions. The functions a_+ and a_- are translated into others of the set by any translation, and hence have zero character for translation. Two of them are left invariant (for a_+) or have their sign changed (a_-). The other kind of reflection interchanges members of the set. The effect of a translation on any u_{p0} is to multiply it by $\exp(ip_0\cdot l/\hbar)$. This number becomes a diagonal element of the matrix representing the translation; the sum of these diagonal elements is the character, and this sum vanishes. Any reflection in points equivalent to the origin interchanges functions of the sets u_{p0} contributing nothing to the character, except for the two band-edge functions. These are left invariant [for u_{p0}(cos)], contributing two, or change sign [for u_{p0}(sin)], contributing minus two to the character. A similar argument shows that the contribution to the character for the other type of reflection and for the sets u_{p0} vanishes for all but the band-edge functions, and the upper and lower band-edge functions give equal and opposite diagonal elements, so that the total is zero.

One kind of set of u_{p0} which has apparently been neglected is that with one cosine-like and one sine-like

[7] G. F. Koster, Phys. Rev. 89, 67 (1952).
[8] P. Feuer, Phys. Rev. 88, 92 (1952).

band-edge function. Actually, this kind of set is included already if we reverse the roles of the two types of reflection, which corresponds to moving the origin from the initial position to the mid-point of the first unit cell. This kind of set could have a sine-like upper-edge function and cosine-like lower one, or *vice versa*. The four possibilities correspond to the four possibilities for band-edge functions discussed originally by Shockley[9] (see Feuer[3]).

The appropriate kind of Wannier functions to go with each of these kinds of bands is easily determined by comparing characters. We note the fact which we have already used, that a_+ is even with respect to reflection at its center, and a_- is odd. Then the set a_+ goes with u_{p_0}(cos) and a_- with u_{p_0}(sin), as these sets have the same characters. Similarly, if we have cos-sin or sin-cos bands, the appropriate Wannier functions would be a_+-like or a_--like, centered at points translationally equivalent to $\frac{1}{2}a$.

If we are willing to give up the ability of the Wannier functions to produce the band-edge eigenfunctions when substituted into Eq. (12) requiring only that u_{p_0} with p_0 different from zero or $h/2a$ be properly rendered, then we have no symmetry requirement at all on the Wannier function. They can be either odd or even, or a linear combination of odd and even functions about the center of localization. An example of this sort of thing occurred when we were able to express a_- in terms of a_+—only the band-edge function gave difficulty in that case. The result is trivial when we recognize that the functions of a band have no symmetry properties which distinguish them from other bands except for the symmetry properties of the band-edge functions. However, it points the way to an understanding of similar but more complicated situations which arise in two- and three-dimensional cases.

In two and three dimensions, situations occur which make it difficult to assign energies to bands—there is a sticking together of bands so that, in effect, one energy may belong to several bands. We wish to settle what restrictions, if any, there may be on the symmetry behavior of a Wannier function of a band of energy levels in a multidimensional crystal. The general case would be rather complicated to carry through, but the kind of result to be expected will be clear from the discussion of a special case, already treated from a somewhat different point of view by Slater[6] and Koster.[7] This special case is a square lattice with a potential given by the sum of Mathieu potentials in x and y, plus a small perturbation. The 0−1 and 1−0 bands are degenerate, and mix under the perturbation to give us a situation like that mentioned in the last paragraph, in which the bands stick together at the center and corners of the Brillouin zone.

It is easy to see that in this case one may choose to represent this pair of bands a pair of Wannier functions which belong to a degenerate representation of the group C_{4v} of the "atomic" sites, since these are in fact the functions with which the analysis of Slater and Koster began. These degenerate functions are not correlated with the individual bands; it takes both Wannier functions to express an eigenfunction of either band. Slater and Koster go on to show how nondegenerate Wannier functions can be formed, each of which will yield by itself, under proper operations, the functions of one band. The point which is not clear so far is how on the one hand a pair of degenerate functions, and on the other a nondegenerate pair, can give the same bands of eigenfunctions. The resolution of the difficulty lies in perceiving that they really do not. Consideration of Slater's[6] Eq. (51) and Table VI indicates that just at the corners and center of the Brillouin zone the nondegenerate Wannier functions produce not the appropriate eigenfunction, but zero. We have a situation like that which we have seen already in the one-dimensional case, when the result of using a Wannier function of the "wrong" symmetry is to have the expression for an eigenfunction at some special $\mathbf{p}_0$ yield zero. This circumstance is also reminiscent of the redundant coordinates which appear in the theory of molecular vibrations. Our conclusion is that we can indeed have nondegenerate Wannier functions if we are willing to give up their ability to reproduce the sticking-point eigenfunctions. This is very much in accord with the considerations of Bouckaert, Smoluchowski, and Wigner,[10] who feel that the definition of a band of energy levels should exclude the sticking-point levels.

Slater's and Koster's nondegenerate Wannier functions do, however, retain the ability to yield the eigenfunction (at least to within a nonzero factor) at points reciprocal space lying on symmetry planes or axes. If this property is given up, then there would be no symmetry restriction at all on the Wannier functions.

TABLE I. Some characters for reducible representations of the linear translation group with reflection.

Basis function	$\{E\|0\}$	$\{E\|na\}$	even n $\{\sigma\|na\}$	odd n $\{\sigma\|na\}$
a_+	N	0	2	0
a_-	N	0	−2	0
u_{p_0}(cos-like)	N	0	2	0
u_{p_0}(sin-like)	N	0	−2	0

RELATION BETWEEN SITE SYMMETRY AND SPACE GROUP SYMMETRY

We intend to show some of the relations between crystal orbital theory and site orbital theory which arise from symmetry and which are of great value in transferring ideas from one theory to the other. The argument can be put in a form applicable to either molecules or crystals; it represents an improvement of

[9] W. Shockley, Phys. Rev. **56**, 317 (1939).

[10] Bouckaert, Smoluchowski, and Wigner, Phys. Rev. **50**, 58 (1936).

a method due to Hall[11] in that here account is taken on the representation of the site group to which the SO belongs. We consider a set of SO, each belonging to the representation γ_j of its site group. If the representation is degenerate we include all the partners. We seek to find which irreducible representations Γ_i of the molecular or crystal (space) group span the reducible representation for which our SO form a basis. We proceed by finding the character $\Xi(R)$ of this reducible representation for each operation of the molecular or crystal group. For a given operation there is a contribution of $\chi_j^{s\cdot g\cdot}(R)$ for each site left invariant by the operation, where $\chi_j^{s\cdot g\cdot}(R)$ is the character for R in the jth site group representation; otherwise the contribution is zero.

$$\Xi(R)=\sum_\alpha \delta_\alpha(R)\chi_j^{s\cdot g\cdot}(R). \quad \delta_\alpha=1 \text{ if } R \text{ is in site group of the } \alpha\text{th site; } \delta_\alpha=0 \text{ otherwise.} \tag{34}$$

We then reduce this representation to find which irreducible representations contribute to it. If $\Xi_i(R)$ is the character of R in Γ_i, we have for the number of times Γ_i appears in the reducible representation

$$^in_i=(1/H)\sum_R \Xi_i^*(R)\Xi(R), \tag{35}$$

where H is the order of the molecular or space group. We substitute for $\Xi(R)$ from Eq. (34) and use the relationship (36), based on the site group's being a subgroup of the molecular or crystal group, to obtain Eq. (37). As Eq. (36b) shows, the a_{ik} are easy to determine once the character tables of the two groups are known:

$$\left.\begin{aligned}\Xi_i(R)&=\sum_k a_{ik}\chi_j^{s\cdot g\cdot}(R), &(36a)\\ a_{ik}&=h^{-1}\sum_R \chi_k^{*s\cdot g\cdot}(R)\Xi_i(R), &(36b)\end{aligned}\right\} R \text{ in site group.}$$

$$^in_i=H^{-1}\sum_\alpha\sum_R\sum_k a_{ik}\chi_k^{*s\cdot g\cdot}(R)\chi_j^{s\cdot g\cdot}(R)\delta_\alpha(R), \tag{37}$$

$$^in_i=a_{ij}. \tag{38}$$

Equation (38) arises from application of the orthogonality relations for group characters and the definition of $\delta_\alpha(R)$ to get for the summation over R and k the value of $a_{ij}h$, where h is the order of the site group and equal to H divided by the number of equivalent sites in the set. Summing over α leaves Eq. (38). This group theoretical argument has been used earlier by the author[12] for similar purposes. Equation (38) tells us which irreducible representations of the crystal or molecular group are included in the representations induced by SO belonging to the jth site group representations located on the sites α, and hence provides a way to translate the symmetry properties of a site treatment into those of a molecular or crystal orbital treatment. As an example of the application of (38) we may determine the number of times any F-dimensional representation of the space group (F is the order of its factor group) appears in the space group representation induced by a set of SO centered on sites of symmetry C_1. From Eq. (36b) a_{ij}, and hence ^in_i, is obviously $(1/1)\chi_1^{C_1}(E)\Xi_i(E)$, or just F.

A converse problem is the determination of the kind of SO which may be constructed if we know that we have molecular or crystal orbitals belonging to Γ_i. In crystals the problem will be nontrivial only when Γ_i is a representation for some special value of $\mathbf{p}_0$ at which the star is degenerate. Possible candidates for SO must be those belonging to such representations γ_j that a_{ij} does not vanish; we can easily determine these by running through all representations of all kinds of sites. Whether the SO going with such j can actually be constructed is further determined by whether we have molecular or crystal orbitals belonging not only to Γ_i but also to Γ_k for all nonzero a_{kj}. Functions of Γ_i need the cooperation of functions in the other representations Γ_k in order that linear combinations of them belonging to γ_j may be set up.

A point worth mentioning is that sites are not necessarily only the loci of atoms in a molecule or crystal. Any point in space is a site with a site group of at least C_1 symmetry. Of course, the more symmetry there is around a site, the finer is the symmetry classification of functions centered at it.

The kind of symmetry information which is given us by the knowledge that we have a set of SO of a certain site symmetry on a set of sites is quite similar, we see from Eq. (38), to information about the symmetry of molecular orbitals or, in crystals, to information about the representations of band functions at special points in crystal momentum space. Further, the SO information is so much more closely related to chemical notions that it may be much more valuable to have the symmetry information in the SO form. In the diamond lattice, for instance, it is much more revealing to know that the valence band functions can be represented by SO (here they are also EO) which are symmetric with respect to inversion at the inversion centers midway between atoms and, hence, are of a binding character than it is to know the complete reduction of the space group representation of these orbitals. Likewise, valuable insight into the conduction band orbitals can be obtained from the "chemical" kind of information that the corresponding SO are antibonding, having a node between atoms required by their antisymmetry under inversion.

EQUIVALENT ORBITALS IN METALS

It seems worth while to see what can be said about metals. The distinguishing characteristic of metals in the band theory is that they do not have filled bands and, hence, it is difficult to define EO for them. However, the artifice of introducing a superlattice, which in turn divides up the Brillouin zones for the original

[11] G. G. Hall, Proc. Roy. Soc. (London) **A202**, 323 (1950).
[12] H. Winston, J. Chem. Phys. **19**, 156 (1951).

lattice, always makes it possible just to include all the occupied crystal orbitals in filled zones of the superlattice. The idea is applicable to ferromagnetic as well as to ordinary metals to the extent that the ground states can be represented by single antisymmetrized determinants. Whether the artifice has any value for quantitative investigations depends on how big the superlattice has to be made; evidently a superlattice so large that one cell of it encompassed the whole crystal would have its zones filled, but it would not simplify things at all. A point of qualitative interest is that the superlattice can be superimposed upon the original lattice in a number of ways equal to the number of original unit cells in one supercell. Going with each of these ways are sets of sets of EO. As Lennard-Jones[2] has shown, the possibility of defining sets of EO in more than one way (as for the π orbitals of benzene) arises in situations where other valence theories would invoke the concept of resonance. Since each of the EO so defined will be spread over many atoms (at least the whole super-cell), we seem to have a parallel to Pauling's ideas[13] on the electronic structure of metals, which are based on an entirely different staring point.

SITE ORBITALS AND PERTURBED LATTICES

SO have a special value in the theory of perturbed lattices. The methods of Peckar[14] and Slater,[1] which use the crystal eigenfunctions or the Wannier functions of one band of the unperturbed lattice as basis functions for a perturbation calculation, are legitimate as long as the perturbation is small compared to the energy separation between bands. This restriction is not satisfied in many practical cases in which degenerate bands occur or where the perturbation is comparable to the separation between nondegenerate bands. Adams[15] has provided a formalism for taking into account as many bands of the unperturbed lattice as may be necessary to expand the perturbed eigenfunctions. We propose a restatement of his method in terms of SO, which has the advantage that SO are generally much easier to visualize than Wannier functions in three dimensions. As in general the SO are linear combinations of Wannier functions, no more than a linear transformation of Adams' equation is necessary to put it in the SO form.

We can discuss qualitatively the perturbation caused by the removal of an atom from a diamond lattice, leaving a vacancy or Schottky defect. The perturbation is very large, compared even to the valence-conduction edge separation. A first approximation would consist of the use of the valence band SO's, localized functions which are even with respect to inversion at the inversion centers of the crystal and which correspond to the chemical idea of bonds. We see immediately, though, that limiting ourselves to expansions in these SO is found to be unsatisfactory, because the complete wave function will be unable to take full account of changes within an atomic distance of the vacancy, limited as it is by the form of the SO. In this case it seems profitable to supplement our basis functions by the SO for the conduction bands, which are odd under inversion. Now the exigencies of the perturbation can cause variations in the relative amounts of odd and even orbital on the sites surrounding the vacancy, and the result is likely to be much more realistic. If convenient, the odd and even SO could be combined to form one set of new SO, placed at points along the interatomic axes not on the inversion center but symmetrically located with respect to it. There will be four such SO nearer each atom than any other. They will be rather like the directed orbitals of a hypothetical carbon atom in space and will be well suited to describing the situation at vacancy, where we may well expect the electronic structure at the neighboring atoms to revert to something like a free-space arrangement.

SPECULATION ON STRUCTURE OF CdSb

Semiconductors are characterized by possessing filled bands, with not too large an energy gap before the next empty band. It is not always easy to see why the energy gaps should come as they do. However, to the extent that we can characterize bonds by the corresponding EO, we can make sense of a filled band structure if it corresponds to a set of filled bonding EO, and if the next SO to be filled are obviously of higher energy. This notion is implicit in the work of Hall on the diamond structure.

The crystal structure of CdSb and ZnSb has been determined by Almin.[16] These semiconducting compounds contain sixteen atoms per unit cell, arranged so that there are twenty-eight interatomic distances corresponding to Sb—Sb bonds and Cd—Sb (or Zn—Sb) bonds. There are also four interatomic distances corresponding to Cd—Cd (or Zn—Zn) bonds. The Cd—Cd, but not the Zn—Zn, distance is significantly larger than can be accounted for by Pauling's atomic radii, according to Almin. We thus expect that the Cd—Cd bonds would have higher energies than the other twenty-eight bonds, and we observe that these other twenty-eight bonds are just filled by the fifty-six valence electrons of the sixteen atoms in the unit cell. The Zn—Zn distance is not indicative of a higher than usual Zn—Zn bond energy, but we can still argue that the Zn—Zn bond is different from the other twenty-eight bonds and very likely has a higher energy even at its normal distance. In each case, then, we can qualitatively assign all the electrons to a set of low-lying bonds, leaving the next higher ones empty. This illustrates the maxim that filled bands imply filled bonds, and vice versa.

[13] L. Pauling, Nature **161**, 1019 (1948); Physica **15**, 23 (1949).
[14] S. Peckar, J. Phys. (U.S.S.R.) **10**, 431 (1946).
[15] E. N. Adams II, Phys. Rev. **85**, 41 (1952).
[16] K. E. Almin, Acta Chem. Scand. **2**, 400 (1948).

APPENDIX

We find directly the relation between $a_{\pm}{}^{l\Delta}(q-Q_n-\Delta)$ and $a_{\pm}{}^{l}(q-Q_k)$ for the nearly-free electron case by writing an expression for the former and expanding it in a series of the latter. For simplicity we will write the a's as $a^{\Delta}(q-Q_n-\Delta)$ and $a(q-Q_k)$, letting the l and $\pm$ be understood. Equation (A1) is derived in the same way as Eq. (9).

$$a^{\Delta}(q-Q_n-\Delta)=(-)^{l-1}\alpha^{-\frac{1}{2}}\frac{\sin\{\pi[l(q-Q_n)-\Delta]/\alpha\}-\sin[\pi(l-1)(q-Q_n)/\alpha]}{\pi(q-Q_n-\Delta)/\alpha}. \tag{A1}$$

Now, $a^{\Delta}(q-Q_n-\Delta)$ can be expanded in series of the orthonormal set $a(q-Q_k)$ for the same band. Determination of the expansion coefficients involves evaluation of the integrals

$$\int dq a^{\Delta}(q-Q_n-\Delta)a(q-Q_k),$$

a long-winded but straightforward process. The result for this coefficient is

$$(-)^{l-1}\frac{\sin\{\pi[l(Q_k-Q_n)-\Delta]/\alpha\}-\sin[\pi(l-1)(Q_k-Q_n)/\alpha]}{\pi(Q_k-Q_n-\Delta)/\alpha}. \tag{A2}$$

A little trigonometric manipulation shows that Eq. (A2) is in agreement with Eq. (26). Writing it in this form, however, brings out the fact that the coefficient is, to within a factor of $\alpha^{-\frac{1}{2}}$, the same function of (Q_k-Q_n) that a^{Δ} is of $(q-Q_n)$ in Eq. (A1).

If we define $a^{\Delta'}(q-Q_n-\Delta)$ as the free electron Wannier function with the phases so chosen that all eigenfunctions are positive real at the point Δ, then obviously $a^{\Delta'}(q-Q_n-\Delta)$ is the same function of $(q-Q_n-\Delta)$ that a $(q-Q_n)$ is of $(q-Q_n)$.

$$a^{\Delta'}(q-Q_n-\Delta)=\alpha^{-\frac{1}{2}}\frac{\sin[l\pi(q-Q_n-\Delta)/\alpha]-\sin[(l-1)\pi(q-Q_n-\Delta)/\alpha]}{\pi(q-Q_n-\Delta)/\alpha}. \tag{A3}$$

As before, the coefficients in the expansion of $a^{\Delta'}$ in terms of the a's can be readily evaluated. The coefficient of $a(q-Q_k)$ in the expansion of $a^{\Delta'}(q-Q_n-\Delta)$ is

$$\frac{\sin[l\pi(Q_k-Q_n-\Delta)/\alpha]-\sin[(l-1)\pi(Q_k-Q_n-\Delta)/\alpha]}{\pi(Q_k-Q_n-\Delta)/\alpha}. \tag{A4}$$

Again, (A4) is of the same form, except for the factor $\alpha^{-\frac{1}{2}}$, as $a^{\Delta'}$ itself in (A3).

The orthogonality properties of the nearly-free-electron Wannier functions can be worked out in terms of the integral

$$g(l,s;m,n)\equiv\int a_{+}{}^{l}(q-Q_m)a_{+}{}^{s}(q-Q_n)dq. \tag{A5}$$

Upon letting $q-Q_m=x$, and $q-Q_n=q-Q_m+(Q_m-Q_n)$ $=x+(m-n)\alpha$, and defining $F(l,s;m-n)$ as

$$\frac{1}{\alpha}\int_{-\infty}^{\infty}dx\frac{[\sin(\pi lx/\alpha)]\sin\{\pi s[x+(m-n)\alpha]/\alpha\}}{\pi^2x[x+(m-n)\alpha]/\alpha^2},$$

we have

$$g(l,s;m,n)=F(l,s;m-n)+F(l-1,s-1;m-n)-F(l,s-1;m-n)-F(l-1,s;m-n). \tag{A6}$$

We can evaluate the F function as follows:

$$F(l,s;r)=[\sin(\langle l,s\rangle r\pi]/r\pi, \tag{A7}$$

where $\langle l,s\rangle$ denotes the smaller of l and s; we get, for $l\leqslant s-1$,

$$g(l,s;m,n)=0, \quad \text{independent of } m \text{ and } n. \tag{A8}$$

For $l=s$ we have

$$g(l,l;m-n)=\delta_{m,n}. \tag{A9}$$

The preceding equations imply $g(l,s;m,n)=\delta_{l,s}\delta_{m,n}$, as required for orthonormality.

Electrical Properties of Pure Tellurium and Tellurium-Selenium Alloys*

ALLEN NUSSBAUM†
Randal Morgan Laboratory of Physics, University of Pennsylvania, Philadelphia, Pennsylvania
(Received November 30, 1953)

The resistivity and Hall coefficient of pure, single-crystals tellurium and of six single-crystal tellurium-selenium alloys, ranging from 2.7 percent to 13.2 percent selenium by weight, were measured over the temperature range 90°K to 550°K, and a few of the samples were taken down to liquid helium temperatures. It was found that the alloys show the same double reversal in the Hall coefficient that is exhibited by pure tellurium, going from *p* type to *n* type in the general region of room temperature, and reversing back to *p* type at temperatures near 200°C. The upper reversal temperature decreases and the forbidden band width, as obtained from the slope of the resistivity *vs* temperature curves, increases with increasing selenium concentration. The room temperature resistivity and lower reversal temperature do not vary monotonically with alloy concentration but a correlation among them has been found based on ideal semiconductor theory. An explanation of the anomalous Hall coefficient which considers the effect of the electron configuration and the lattice structure on the energy band picture is discussed and it is shown how this hypothesis fits the observed facts. As a further test of this theory, the effect of a hydrostatic pressure of 5000 lb/in. on the two Hall reversal temperatures was measured.

I. INTRODUCTION

THE electrical properties of single-crystal tellurium of various degrees of purity have recently been studied extensively at Purdue University[1] and at Tôhoku University[2,3] in Japan. The results of this work show that very pure tellurium is a *p*-type semiconductor below about −40°C, that it reverses to *n* type at or above this temperature, depending on the impurity concentration, and that it reverses back to *p* type at about 230°C, the latter temperature being independent of concentration. The forbidden band width, as obtained from the slope of the resistivity *vs* temperature curve, is 0.32 ev. In the work to be described here, we have extended these electrical measurements to tellurium-selenium alloys, the maximum selenium concentration being about 15 percent by weight. The anomalous second reversal was found in all samples studied, and since tellurium, selenium, and their alloys in all concentrations are isomorphous,[4] this phenomenon may be associated with the hexagonal lattice structure of these materials, the lattice being formed from spiral chains lying side-by-side.[5] Callen[6] and Miller[7] have proposed an explanation of the second reversal to be described below, which is based on a qualitative description of the variation of the energy bands with lattice spacings. This energy band picture is believed to arise from the combination of the group VIb electronic configuration s^2p^4 and the unusual lattice structure. In order to verify some of the predictions of the Callen-Miller scheme, the effect of an applied hydrostatic pressure of 5000 lb/in.[2] on the two reversal temperatures of some 15 percent alloys was measured. It was found that this band picture also fits our other experimental results, namely an increase in the forbidden band width and a lowering of the anomalous reversal temperature with increased selenium concentration.

II. EXPERIMENTAL PROCEDURES

The resistivity and Hall coefficient of a sample of pure single-crystal tellurium and of six single-crystal tellurium-selenium alloys were measured as a function of temperature over the range of 90°K to 550°K, and a few of the samples were carried down to liquid helium temperatures. These samples were prepared by Loferski, formerly of this laboratory, who used a modification of the Bridgman method.[8] The samples were cylindrical in shape, with a length of about 2 cm and a diameter of about 1 mm. Leads for electrical measurements consisted of 20-mil platinum wire welded to the sample by passing a high current through the wire. The temperature range of room temperature to liquid air (or liquid helium) was obtained by use of a set of glass double dewars. The sample was in good thermal contact with a 660-g copper block and a platinum resistance thermometer was used for temperature measurements. For the high-temperature range, the sample, a thermocouple, a stirring rod, and a nichrome heater were immersed in a beaker of Prestone oil. This oil is actually a polyalkylene glycol and its charring point is about 275°C. Conventional dc methods were used to measure the resistivity and Hall coefficients, voltages being measured with either a Rubicon Model 2780 Potentiometer or a Leeds and Northrup Model 9835A Indicating dc amplifier.

* Supported by a U. S. Air Force Contract.
† Now at General Research Laboratories, Minneapolis-Honeywell Regulator Company, Minneapolis 8, Minnesota. Abstracted from part of a thesis submitted in partial fulfillment of the requirement for the Ph.D. in Physics at the University of Pennsylvania, October, 1953.
[1] V. E. Bottom, Science **115**, 570 (1952).
[2] Fukuroi, Tanuma, and Tobisawa, Science Repts. Research Insts. Tôhoku Univ. **A1**, 373 (1949).
[3] T. Fukuroi and S. Tanuma, Science Repts. Research Insts. Tôhoku Univ. **A4**, 353 (1952).
[4] E. Grison, J. Chem. Phys. **19**, 1109 (1951).
[5] A. von Hippel, J. Chem. Phys. **16**, 372 (1948).
[6] H. B. Callen, J. Chem. Phys. **22**, 518 (1954).
[7] P. H. Miller, Jr. (private communication).
[8] J. J. Loferski, Phys. Rev. **93**, 707 (1954).

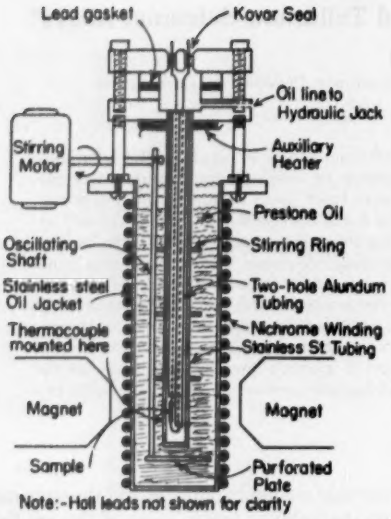

FIG. 1. Apparatus for measuring Hall coefficient under applied pressure.

The apparatus for the measurement of the effect of hydrostatic pressure on the Hall coefficient is shown in Fig. 1. The hydraulic fluid was Prestone oil and precautions were taken to insure thermal equilibrium and to eliminate errors in temperature measurements. The auxiliary heater at the top of the bomb was maintained at a temperature a few degrees above that at the bottom in order to establish a thermal block and thus reduce the internal circulation of the oil. The alloys containing 15 percent selenium were selected for study in this apparatus, since these samples have the lowest anomalous reversal temperature and hence were the easiest to study. *Note added in proof.*—All of the single crystals and their C axis within 10° of the cylinder axis.

III. EXPERIMENTAL RESULTS

The electrical characteristics of the seven samples which were studied are tabulated in Table I.

The resistivity and Hall coefficient curves for the 7.1 percent sample are shown in Figs. 2 and 3, respectively, and the curves for the other samples appear in a separate report.[9] In addition to the pure tellurium sample shown in Table I, a run over a narrower temperature range was made on a sample of cp grade, which was prepared by recrystallizing the material as furnished by the supplier. Bottom[1] has shown that the purer the tellurium, the lower is the temperature at which the material reverses from p type to n type, and the results for our two tellurium samples were found to interlock very well with those for four of his samples.

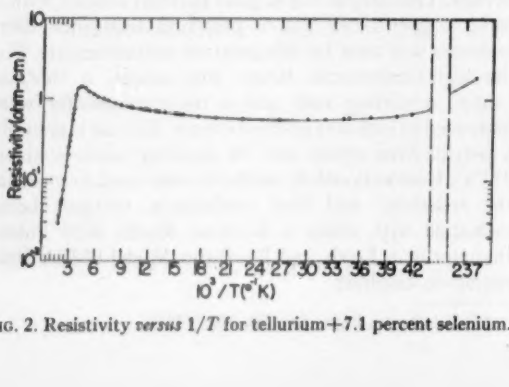

FIG. 2. Resistivity *versus* $1/T$ for tellurium+7.1 percent selenium.

The effect of applied hydrostatic pressure on the Hall coefficients of two different alloys each containing about 15 percent selenium is shown in Figs. 4 and 5, and it is seen that both the upper and lower Hall reversal temperatures are reduced by pressure.

IV. DISCUSSION

A. Correlation of Results with Alloy Composition

The experimental results tabulated in the previous section indicate that some of the electrical properties of our samples change monotonically as we proceed from pure tellurium to the alloy with the highest selenium concentration, whereas others fluctuate in a seemingly erratic fashion. In order to correlate these latter parameters, we will derive an expression connecting the lower reversal temperature T_r with the intrinsic conductivity of an ideal semiconductor consisting of a filled band, a conduction band, and an acceptor level located quite close to the filled band. By definition:

$$\sigma = n_e e\mu_e + n_h e\mu_h = \mu_h e(n_e c + n_h), \quad (1)$$

where σ = conductivity, n_e = density of free electrons in conduction band, n_h = density of free holes in filled band, μ_e = electron mobility, μ_h = hole mobility, $c = \mu_e/\mu_h$, and e = absolute value of the charge on the electron. The intrinsic concentration can be defined by the equation:

$$n_i^2 = n_e n_h. \quad (2)$$

TABLE I. Summary of resistivity and Hall data for tellurium and tellurium-selenium alloys.

Composition by weight	Intrinsic resistivity at 23°C (ohm-cm)	Forbidden band width (ev)	Lower Hall reversal temperature (deg C)	Upper Hall reversal temperature (deg C)	Saturation Hall coefficient (cm³/coulomb)
Pure Te	0.54	0.320	−43°	227°	3.8 ×10³
Te+2.7% Se	0.96	0.336	60°	220°	2.8 ×10³
Te+5.7% Se	1.24	0.360	50°	215°	8.5 ×10³
Te+7.1% Se	0.76	0.370	13°	214°	4.3 ×10³
Te+10.9% Se	7.40	0.382	28°	203°	3.2 ×10³
Te+11.3% Se	9.50*	0.400	112°	203°	6.6 ×10³
Te+13.2% Se	5.40*	0.400	49°	177°	1.52×10³

* Extrapolated.

[9] A. Nussbaum, Technical Report No. 1, Contract AF-33-(616)-78 (unpublished). (Copies may be obtained from the Randal Morgan Laboratory of Physics, University of Pennsylvania, Philadelphia 4, Pennsylvania.)

It is known that the intrinsic concentration n_i can be also expressed as[10]

$$n_i=(PQ)^{\frac{1}{2}}\exp(-E_G/2kT),$$

where

$$P=2(2\pi m^* kT)^{\frac{3}{2}}/h^3,$$
$$Q=2(2\pi m^{**} kT)^{\frac{3}{2}}/h^3, \qquad (3)$$

m^*=effective mass of an electron, m^{**}=effective mass of a hole, T=absolute temperature, h=Planck's constant, k=Boltzmann's constant, and E_G=forbidden band width. Combining (1), (2), and (3) gives an equation which is valid at intrinsic temperatures, and in particular at the lower Hall reversal temperature T_r. Incorporating with this result the condition for the Hall coefficient reversal,

$$n_r c^2=n_h,$$

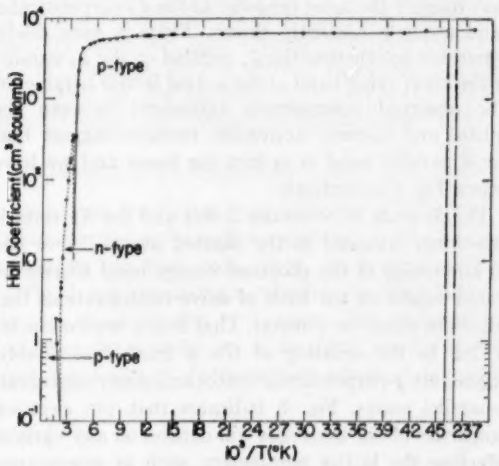

FIG. 3. Hall coefficient *versus* $1/T$ for tellurium +7.1 percent selenium.

gives

$$\sigma_r=\mu_{hr}e(c+1)(PQ)^{\frac{1}{2}}\exp(-E_G/2kT), \qquad (4)$$

where the subscript r refers to values at T_r. The values of μ_{hr} to be used in this equation are obtained by computing the Hall mobility in the extrinsic range from the equation:

$$\mu_h=3\pi R/8\rho, \qquad (5)$$

where R is the Hall coefficient and ρ the resistivity. We then plot mobility *vs* temperature and determine μ_{hr} by extrapolating the curve into the intrinsic region, assuming a $T^{-\frac{3}{2}}$ law. This curve for one sample is shown in Fig. 6. The values of (c) to be used in (4) may be estimated in several ways.[11–13] The coefficient of the exponential will not be an explicit function of temperature since the $T^{\frac{3}{2}}$ term contained in $(PQ)^{\frac{1}{2}}$ will be cancelled out by the $T^{-\frac{3}{2}}$ factor in the mobility. Hence, if we plot $\{\ln\sigma_r/\mu_{hr}(c+1)\}$ *vs* E_G/T_r, we should get a straight line of slope $2k$. Using the tabulated data, we get the results shown in Fig. 7, which indicates that there is a correlation of the various electrical properties with alloy concentration.

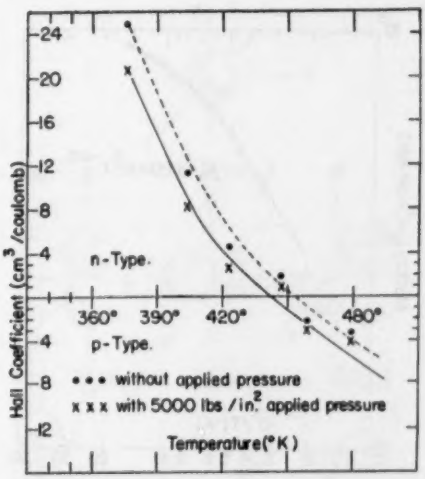

FIG. 4. Shift of upper reversal temperature with applied pressure for tellurium+15 percent selenium.

B. Energy Band Picture for Tellurium and Selenium and Its Consequences

The original attempt to explain the second Hall reversal in tellurium on the basis of its band structure was made by Fukuroi and Tanuma.[3] They postulated the existence of two conduction bands, overlapping in such a manner that the bottom of one band lies somewhat higher than the other. They then say that the

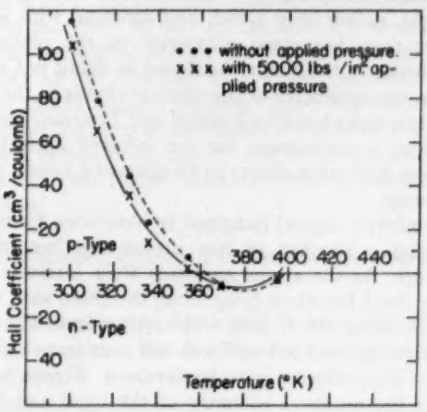

FIG. 5. Effect of applied pressure on low-temperature Hall reversal for tellurium+15 percent selenium.

[10] Hunter, Rittner, and duPre, Philips Research Repts. **5**, 188 (1950).
[11] L. P. Hunter, Phys. Rev. **91**, 207 (1953).
[12] A. Nussbaum, Technical Report No. 8, Department of Physics, University of Pennsylvania, 1953 (unpublished).
[13] G. L. Pearson and J. Bardeen, Phys. Rev. **75**, 865 (1949).

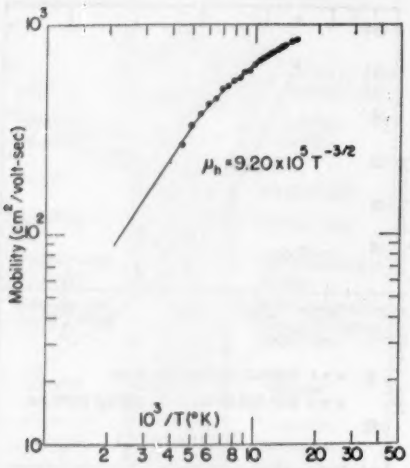

FIG. 6. Hole mobility *versus* $1/T$ for tellurium plus 7.1 percent selenium.

effective mass of a free electron will be the average of the effective masses in each of the two bands, and since the relative concentration of electrons in the two bands will depend on temperature, the average effective mass will also do so. Thus, the mobility ratio will also vary with temperature, decreasing when the ratio of effective masses increases, and it presumably approaches unity as the temperature rises to 230°C. The Hall coefficient will then go from negative to positive in accordance with

$$R = -\frac{3\pi(c^2-1)}{8en_i(c+1)^2}. \tag{6}$$

Callen[6] has suggested that a number of properties of tellurium and selenium may be understood on the basis of a particular tetragonal simplification of the true crystal structure. He considers a hypothetical structure in which the chains of the tellurium structure are straight, rather than spiral, and arranged with square rather than hexagonal symmetry in the plane perpendicular to their axes. As shown in detail in Callen's paper, the symmetry of the structure leads to the band structure postulated by Fukuroi and Tanuma,[3] and also provides a mechanism for the infrared optical phenomena and other effects to be discussed briefly in the following.

Consider a crystal imagined to have been formed by bringing a number of free atoms together from a distance. As the atoms approach their ultimate positions, the *d* functions lying along the chain axis, which is denoted as the *C* axis, will overlap first. The wave functions parallel to the *C* axis will overlap to a greater extent than those perpendicular to it. Figure 8 illustrates the resultant widening of the band and Fig. 9 shows the symmetry properties of the functions in the isolated atoms. In Fig. 8, we have plotted reduced, rather than actual, lattice parameters. These are defined as the ratio of the lattice parameters along each axis at a given temperature and pressure to their values at an arbitrary temperature and pressure. The use of reduced quantities enables us to plot both sets of parameters on one diagram. The next functions to overlap are the *d* functions normal to the chain axis, then the *p* functions, and finally the *s* band widens. The doubly degenerate *p* functions $\psi_p(E_v)$ give rise to the filled band, and the lower-lying *s* band needs not be further considered. For convenience in following the subsequent discussion, we will denote the wave functions corresponding to *p* functions oriented parallel to the chain axis by "*p* parallel," with similar designations for the other functions.

The other three *d* functions which are not shown in either Fig. 8 or Fig. 9 actually give rise to bands which have roughly the same behavior as the ψ_d-perpendicular band which is explicitly shown. There is some doubt *a priori* as to whether the ψ_p parallel or the ψ_d parallel in the lower lying band at the actual lattice parameter. The observed approximate agreement between the optical and thermal activation energies suggest that the *d* parallel band is in fact the lower and we have drawn Fig. 8 accordingly.

The 5*s* state in tellurium is full and the 5*p* state is two-thirds occupied in the isolated atoms. If we had no knowledge of the proposed energy band scheme we would expect on the basis of naive considerations that tellurium would be a metal. That is is a semiconductor is due to the splitting of the *p* levels into doubly-degenerate *p*-perpendicular states and singly-degenerate *p*-parallel states. Fig. 8 indicates that the electrical properties of our materials will depend on any variable affecting the lattice parameters, such as temperature

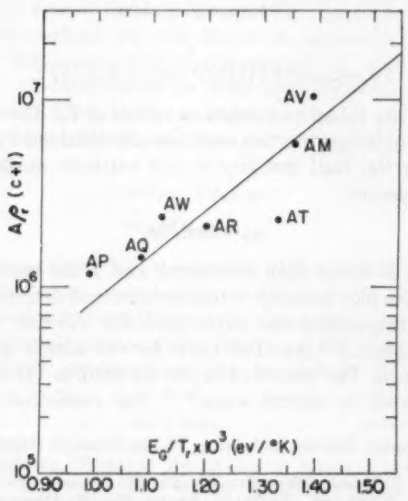

FIG. 7. Correlation of electrical properties with alloy concentration in tellurium-selenium alloys.

and pressure. The following experimental results illustrate this.

1. High-Temperature Reversal

Since the upper reversal in the Hall coefficient occurs at temperatures high enough to assume that the number of free electrons equals the number of free holes, if the mobility ratio in tellurium is reduced to unity, the Hall coefficient will approach zero. In Fig. 8, we assume that at temperatures near the lower Hall reversal temperature, when electrons are just beginning to be thermally excited across the forbidden gap, the majority of them will go into the lower-lying *d*-parallel band. As the temperature increases, a larger fraction of the free electrons get into the higher *d*-perpendicular band. Since this is a narrower band with a lower mobility, the average mobility of the free electrons will decrease with temperature, and eventually the effective mobility ratio will pass through unity to fractional values. Thus we have a second reversal in the Hall coefficient. In this explanation, we have ignored the effects arising from changes in crystal dimensions.

2. Increase of Forbidden Band Width in the Alloys

As shown by Grison,[4] the addition of selenium to tellurium reduces the lattice parameters both along and normal to the chain axis. Thus, we move to the left in Fig. 8 and the distance between the filled and conduction bands is seen to increase. The values of the forbidden band width we have given are determined from the slopes of the resistivity curves in the region

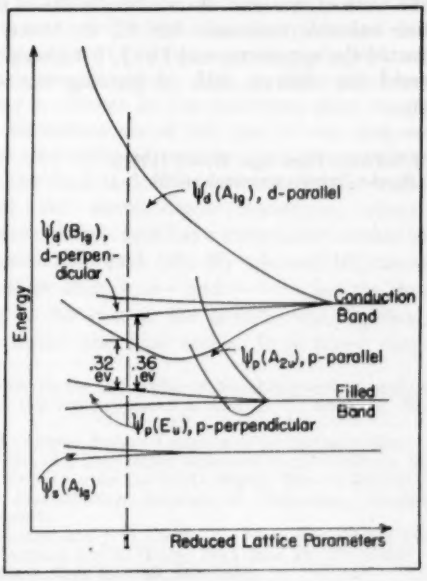

FIG. 8. Schematic dependence of energy bands on lattice parameters.

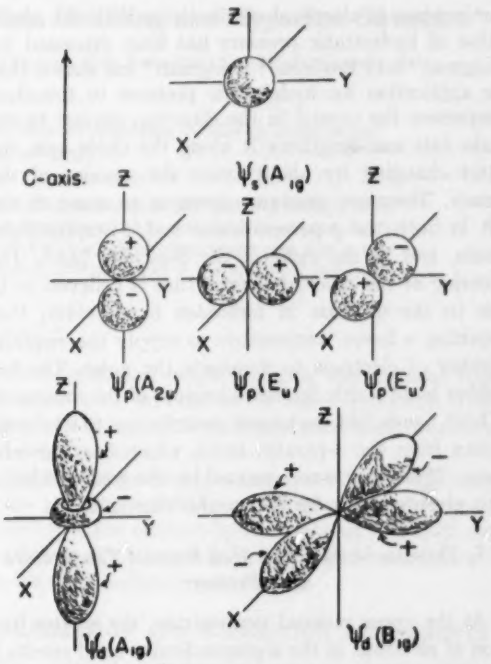

FIG. 9. Symmetry properties of isolated atom wave functions.

below the second reversal points, so that it is correct to take the forbidden gap as the distance between the top of the *p*-perpendicular band and the bottom of the *d*-parallel one. Also, we have made the reasonable assumption that the *d*-parallel band, with its greater overlap of functions, will have the greater slope.

3. Decrease in the Upper Hall Reversal Temperature in the Alloys

As explained in Sec. 1, above, the anomalous reversal occurs because of the greater fraction of electrons being found in the *d*-perpendicular band at high temperatures, and as described in Sec. 2, the addition of selenium reduces the lattice parameters. Hence, if we move to the left in the diagram, at a given temperature we would expect a larger fraction of the free electrons in the upper *d*-perpendicular band, because of the closer spacing between the upper and lower perpendicular bands, and the second reversal occurs at a lower temperature.

4. Decrease in Lower Hall Reversal Temperature with Pressure

At temperatures in the vicinity of room temperature, it is believed that the mobility ratio is essentially independent of temperature, and we have the usual criterion for a reversal:

$$n_e c^2 = n_h.$$

The increase in electrical conductivity with the application of hydrostatic pressure has been discussed by Bridgman[14] and Bardeen.[15] Bridgman[16] has shown that the application for hydrostatic pressure to tellurium compresses the crystal in the direction normal to the chain axis and lengthens it along the chain axis, the latter changing by about twice the amount of the former. Therefore pressure causes us to move to the left in both the *p*-perpendicular and *d*-perpendicular bands, and to the right in the *d*-parallel band. The lowering of the reversal temperature is believed to be due to the decrease in forbidden band width, thus requiring a lower temperature to supply the requisite number of electrons to dominate the holes. The forbidden band width decreases because of the movement of both bands, but the largest contribution to its change comes from the *d*-parallel band, which has a greater slope. This effect is accentuated by the greater dilation at a given pressure for the parallel dimension.

5. Decrease in the Upper Hall Reversal Temperature with Pressure

At the upper reversal temperature, the greater fraction of electrons in the *d*-perpendicular band results in a lowering of the mobility ratio. Applying pressure, it is seen that the distance between the two perpendicular bands lessens and causes the reversal temperature to decrease, in a fashion similar to that described in Sec. 2, above, but the distance between the *p*-perpendicular and *d*-parallel bands increases, with presumably the opposite effect. However, the density of states in the *d*-perpendicular band is believed to be so much higher than the *d* parallel, that the average mobility is reduced, producing a lowering of the reversal temperature.

[14] P. W. Bridgman, Proc. Am. Acad. Arts Sci. **72**, 159 (1938).
[15] J. Bardeen, Phys. Rev. **75**, 1777 (1949).
[16] P. W. Bridgman, Proc. Am. Acad. Arts Sci. **77**, 187 (1949).

6. Optical Properties

The transmission measurements made on single crystals of tellurium by Loferski[17] show that there are two absorption edges, depending on whether the incident light is polarized parallel or perpendicular to the *C* axis. The forbidden band widths of 0.36 ev for the former case and 0.32 ev for the latter are shown in Fig. 8. The selection rules devised by Callen[6] explain this dependence of the absorption edge on the direction of polarization.

C. Alloy Scatterings

Brooks[18] has suggested that in addition to impurity and thermal scattering normally considered in mobility studies, there will exist in alloys a scattering due to the deformation of the band edges produced by the minority component with its different forbidden band width. We have been unable to detect this alloy scattering in our samples because of the relative magnitudes of thermal and alloy mobilities. The total mobility will be the reciprocal of the sum of the reciprocals of the two contributions, and since for our samples the alloy mobilities as predicted by Brooks are about 50 times higher than the computed Hall mobilities, they would have very little effect on the total.

V. ACKNOWLEDGMENTS

The author would like to express his gratitude to Professor P. H. Miller, Jr., who not only suggested this problem and provided constant guidance and technical advice, but also gave much-needed encouragement during the many difficult periods which are an inevitable part of a work of this kind. He would also like to thank for their valuable assistance Mr. H. B. Mohr, who constructed the apparatus, and Dr. J. J. Loferski, who performed the difficult task of growing the single crystals.

[17] J. J. Loferski, Phys. Rev. **87**, 905 (1952).
[18] H. Brooks (private communication).

Microwave Resonance in Nitric Oxide: Lambda Doubling and Hyperfine Structure*

ROBERT BERINGER AND E. B. RAWSON,† *Sloane Physics Laboratory, Yale University, New Haven, Connecticut*

AND

ALLAN F. HENRY, *Atomic Power Division, Westinghouse Electric Corporation, Pittsburgh, Pennsylvania*

(Received November 17, 1953)

The microwave spectrum arising from transitions between Zeeman levels in the $^2\Pi_{\frac{1}{2}}$, $J=\frac{1}{2}$ state of nitric oxide is reinvestigated with particular regard to the lambda doubling which is observable for electric-dipole transitions. This doubling is compared with the theory of Karnaugh. The experiments also give a new and improved determination of the nuclear hyperfine coupling constants and of the molecular splittings. The theory of the hyperfine effect is reexamined in detail.

INTRODUCTION

MICROWAVE studies[1] and the theory[2,3] of the high-field Zeeman levels of the $J=\frac{1}{2}$, $^2\Pi_{\frac{1}{2}}$ state in $N^{14}O^{16}$ have appeared in this journal. After publication, J. H. Van Vleck pointed out to us that each Zeeman level is a lambda doublet, that electric-dipole transitions between the lambda-doublet components of adjacent Zeeman levels would occur, and that the lambda doubling would be observable with sufficient resolution. This paper reports an experimental investigation of these suggestions, a recalculation of the hyperfine structure constants from the improved date, and a further check of the molecular Zeeman theory. A forthcoming paper of Karnaugh[4] discusses the theory of the lambda doubling in the nitric-oxide Zeeman levels.

The microwave spectrum of $N^{14}O^{16}$ in high magnetic fields was originally interpreted[1] as being due to magnetic-dipole transitions (M_J, $M_I \rightarrow M_J \pm 1$, M_I) between the twelve Zeeman levels arising from the lowest ($J=\frac{1}{2}$) rotational level of the $^2\Pi_{\frac{1}{2}}$ component. This interpretation is still correct for absorption and induced emission produced by microwave magnetic fields normal to the dc magnetic field. Also, no appreciable lambda doubling is present in this spectrum, since magnetic-dipole transitions are of the type $+\rightarrow+$ and $-\rightarrow-$ and the alternation of $+$ and $-$ levels serves almost to cancel the lambda-doublet intervals in the transition energies. For electric-dipole transitions, where the microwave electric field has a component normal to the dc magnetic field, each (M_J, $M_I \rightarrow M_J \pm 1$, M_I) transition is a doublet since $+\rightarrow-$ and $-\rightarrow+$, and the doublet spacing is the sum of the lambda-doubling intervals in the initial and final states. It is noted that the electric-dipole and magnetic-dipole spectra arise from the same levels and are of the same frequency. They differ only in regard to the lambda doubling, intensity, and polarization.

The most effective way to separate the two spectra is by the polarization of the microwave fields; the greater intensity of the electric-dipole spectrum is also of use. In addition, the observation of lambda doubling identifies the electric-dipole spectrum.

POLARIZATION AND SATURATION EXPERIMENTS

The original experiments[1] used a circular TM_{011} mode cavity with the cavity axis parallel to the dc magnetic field. In this field configuration the microwave H, being purely azimuthal, is entirely perpendicular to the dc H, but only the radial component of the microwave E is perpendicular to the dc H. The electric-dipole spectral intensity is proportional to the square of this radial E integrated over the cavity volume, and the magnetic-dipole intensity is proportional to the integral of the square of the azimuthal H integrated over the cavity volume. This integration for the cavity used gives an electric-dipole spectrum 34 percent as intense as the magnetic-dipole spectrum for the same intrinsic intensity. The ratio of the intrinsic electric to magnetic intensity[5] is μ_e^2/μ_0^2, where[6] μ_e is the permanent electric-dipole moment and μ_0 is the Bohr magneton. Assuming that $\mu_e = 0.1\times10^{-18}$ esu, the intrinsic intensity ratio is about 120, and in the reported experiments[1] the electric spectrum would have been 40 times as intense as the magnetic spectrum. It appears, therefore, that the agreement cited in reference 1 between the measured intensity and that calculated for magnetic-dipole transitions was quite erroneous. Part of this discrepancy is attributable to power saturation of the electric-dipole spectrum which, for the conditions prevailing in reference 1, would have reduced the electric-dipole spectral intensity by a factor 2. This leaves a factor 20 for the

* Assisted by the U. S. Office of Naval Research. A preliminary report of this work appeared in Ann. N. Y. Acad. Sci. **55**, 814 (1952).

† U. S. Atomic Energy Commission predoctoral fellow. This work is part of a dissertation submitted to the Graduate School of Yale University for the Ph.D. degree. Now at Lincoln Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.

[1] R. Beringer and J. G. Castle, Jr., Phys. Rev. **78**, 581 (1950).
[2] H. Margenau and A. Henry, Phys. Rev. **78**, 587 (1950).
[3] A. Henry, Phys. Rev. **80**, 549 (1950).
[4] M. Karnaugh, Ph.D. dissertation, Yale University, 1952 (unpublished).

[5] Equation (15) of reference 2 gives the magnetic-dipole matrix element. The electric-dipole matrix element has the same dependence on J and M and the same factor 0.386.

[6] Watson, Rao, and Ramaswamy, Proc. Roy. Soc. (London) **143**, 558 (1934), give $\mu_e = 0.16\times10^{-18}$ esu. C. P. Smyth and K. B. McAlpine, J. Chem. Phys. **1**, 60 (1933), give $\mu_e < 0.1\times10^{-18}$ esu.

discrepancy in the ratio of the calculated to observed intensities in reference 1.

The first experiment of the present work used a rectangular TE_{01} mode cavity one-half wavelength long. It was constructed of a copper wave guide of internal dimensions 0.900 in.×0.400 in. and coupled at either end to a similar wave guide by means of inductive irises. The microwave apparatus and experimental method were similar to those described.[1] Microwave radiation at the cavity-resonance frequency was coupled through the cavity to a dc biased bolometer. The cavity was in the gap of a magnet, and the magnetic field was varied to produce resonance between the cavity frequency and the spectral transitions. This resonance diminished the power coupled to the bolometer and was detected by modulating the magnetic field across the spectral line at a low audio-frequency and noting the modulation in the voltage across the bolometer. After amplification and demodulation in a lock-in mixer operating at the modulation frequency, the absorption signals were presented on a galvanometer.

The electromagnet, of recent construction, was operated from a motor-generator controlled by a simple, degenerative voltage stabilizer operating on the shunt field of the generator. A proton-resonance field regulator feeding back to a high impedance winding on one magnet pole was used for accurate field regulation. The 30-cy/sec field modulation was produced by a winding on the other pole.

In the rectangular TE_{011} mode the microwave magnetic field is confined to planes parallel to the broad sides of the wave guide, and the electric field is perpendicular to these planes. If the wave guide axis is perpendicular to the dc magnetic field, rotation of the wave guide about its axis can bring either the microwave electric or magnetic field entirely into planes perpendicular to the dc field. When the magnetic dipole spectrum is being observed polarization entirely forbids the electric dipole transitions, but when the electric spectrum is observed some magnetic spectrum is mixed in. Experiments were tried in which the cavity was rotated as described. The electric dipole spectrum was easily observed but the magnetic dipole spectrum was too weak for detection. The indicated ratio μ_e^2/μ_0^2 from these observations was greater than 20, setting a very loose lower limit of 4×10^{-20} for μ_e.

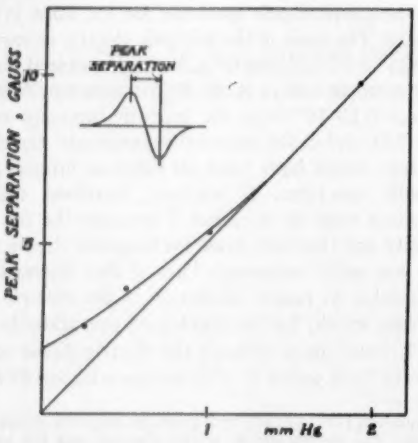

FIG. 1. Variation of width of electric-dipole line H_1 with pressure showing systematic broadening at low pressures. This was evidence for unresolved Λ components spaced by approximately 2 gauss.

Using the same cavity a measurement of μ_e^2 was made from power saturation of the electric spectrum. The energy density in the cavity was fixed and the gas pressure successively reduced to produce increasing saturation. Analysis of the results with the usual theory gave $\mu_e=0.3\times10^{-18}$ esu. This is in disagreement with the static value[6] and if used in calculating absolute intensity would give an even larger discrepancy in the measurements of reference 1 than that cited above. The situation regarding the spectral intensities in the various measurements is quite confused. It is further complicated by the recent work here of R. Collier, who has attempted without success to detect the direct Λ-doubling transitions. From his apparatus sensitivity he concludes that the μ_e is less than 0.06×10^{-18} esu.

In the saturation experiments the lambda doubling was not resolved at the lowest usable pressures, but there was some suggestion of doublet structure from the relatively greater breadth of the line at low pressure. This is shown in Fig. 1 where the line widths are plotted. In these measurements the modulation amplitude was was scaled linearly with the pressure so as to minimize modulation broadening. Extrapolation to zero pressure yields about 2 gauss for the peak separation, which is consistent with a line having two components spaced by about the same amount. This extrapolated component spacing is somewhat high because of the broadening due to saturation.

Analysis of this and similar data also gives the width of the NO lines in the absence of lambda doubling and power saturation. This was found to be $\Delta H=3.0$ gauss/mm Hg using the methods and notation of Fig. 4 in reference 1. This line width corresponds to 3.3 (Mc/sec) per mm Hg,[7] whereas the earlier[1] value was $\Delta H=3.5$ gauss/mm Hg at 9360 Mc/sec or 3.8 (Mc/sec) per mm Hg. The lower line width in the present experiments is due to improved field regulation.

In order to observe the unsaturated lines at pressures low enough to resolve the lambda doubling, it was necessary to modify the apparatus. A lower power level would have sufficed, but with our apparatus the signal-to-noise ratio of the bolometer system suffers severely at low microwave power. In the absence of

[7] It is not quite true that line widths in frequency units are $\Delta\nu=\nu(\Delta H/H)$ because of the nonlinear (Paschen-Back) type of variation of level energy with H. In the present experiment this correction to $\Delta\nu$ is small compared with the experimental error in ΔH.

TABLE I. Experimental lambda-doubling and resonance field values for $N^{14}O$ spectrum.

Line	M_J, $M_I \rightarrow M_J-1$, M_I	Corrected A component separation (gauss)	Arithmetic average resonance field of A components (gauss)	Microwave frequency (Mc/sec)
H_1	$\frac{3}{2},1\rightarrow\frac{1}{2},1$	1.48	8398.61	9269.94
H_2	$\frac{3}{2},0\rightarrow\frac{1}{2},0$	1.34	8425.71	9270.65
H_3	$\frac{3}{2},-1\rightarrow\frac{1}{2},-1$	1.32	8453.44	9270.65
H_4	$\frac{1}{2},1\rightarrow-\frac{1}{2},1$	1.54	8501.49	9270.65
H_5	$\frac{1}{2},0\rightarrow-\frac{1}{2},0$	1.86	8528.28	9270.65
H_6	$\frac{1}{2},-1\rightarrow-\frac{1}{2},-1$	1.49	8555.84	9270.65
H_7	$-\frac{1}{2},1\rightarrow-\frac{3}{2},1$	1.82	8600.97	9269.81$_5$
H_8	$-\frac{1}{2},0\rightarrow-\frac{3}{2},0$	1.97	8629.71	9270.65
H_9	$-\frac{1}{2},-1\rightarrow-\frac{3}{2},-1$	2.01	8656.46	9270.65

sensitive, low-power bolometers, saturation was alleviated by using a large cavity which lowered the microwave field intensity for a given power transmitted to the bolometer. A cylindrical TE_{011} mode cavity of 1.99-in. i.d. was used. The cavity axis was along the dc magnetic field. Another design feature was to make the input coupling much smaller than the output coupling to the bolometer. This serves to keep the loaded Q high, and hence the spectrometer sensitivity large for a given transmitted power. The cavity used had a power transmission of 31 percent, loaded Q of 2850 and unloaded Q of 21 000. At a bolometer power level of 10^{-3} w, a pressure of 0.1 mm Hg in this cavity produced the same degree of saturation as did 0.7 mm Hg in the rectangular TE_{011} cavity. With pressures of 0.1 mm Hg the lambda doubling could be resolved.

The Lambda Doubling

A number of runs were taken with the circular TE_{011} cavity to determine the lambda doubling in the nine absorption lines. The microwave frequency was fixed at the center of the cavity resonance in each run. The dc magnetic field was varied with a proton field control which employed a proton oscillator[8] locked to a variable and accurately known frequency of about 36 Mc/sec provided by mixing harmonics of a BC 221 frequency meter with a 30-Mc/sec crystal-controlled signal. As in earlier work[1] the data were taken point-by-point, varying the dc magnetic field and noting the output of a lock-in mixer with a galvanometer.

The microwave cavity frequency was measured by mixing the microwave oscillator signal with a crystal-controlled 9270-Mc/sec signal and measuring the difference frequency with a calibrated communications receiver. The cavity frequency was fixed throughout a run to about ± 2 kc/sec.

Figure 2 is a typical example of these measurements. The two lambda-doublet components are seen to be resolved well enough to determine their separation. The two outer "crossovers" (points of zero lock-in mixer output) are slightly closer together than the centers of the two line components because of the slope of the overlapping line tails. This "squeezing" correction was evaluated by graphical analysis of the partially resolved patterns. In this analysis it was assumed that the two line components were identical but not necessarily symmetrical. Line contours were added until the experimental line shape was reproduced. Such fitting is very sensitive to the separation of the two line components when they are resolved as well as those of Fig. 2.

Table I gives the results of measurements on the nine lines. The line component separations have been corrected for "squeezing." The fields in the third column are the arithmetic mean fields of the two lambda components and are corrected for field difference between the proton probe and the center of the cavity.

The lambda-doubling theory of Karnaugh[4] expresses the separation of + and − levels in a given (M_I,M_J) state of $N^{14}O^{16}$ in terms of four parameters which must be evaluated by experiment. Two of these arise from the nonhyperfine part of the doubling and the other two from hyperfine contributions. The nine (M_I, $M_J\rightarrow M_I$, M_J-1) transitions between the twelve levels serve to evaluate these four parameters. Karnaugh does this in two stages: the average doubling in each group of three lines with the same M_J, M_J-1 is used to evaluate the nonhyperfine contributions. These three averages are satisfied by the theory with two parameters to within 0.14 percent. The two hyperfine parameters are then evaluated by taking differences in the doubling in several states differing in M_I. Table II gives the result of these calculations. The deviations between the experimental and semitheoretical values of the lambda doubling are well within the experimental error which

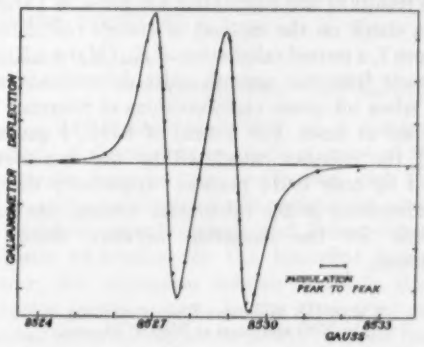

FIG. 2. Signal from line H_5. The gas pressure was 0.096 mm Hg, 30 cy/sec, modulation was 0.68 gauss peak-to-peak, and the microwave frequency 9270.650 Mc/sec. The outer crossovers are separated by 1.84 gauss which is corrected by 0.02 gauss for "squeezing" to give 1.86 gauss for the separation of the A doublet components.

[8] In these experiments a regenerative, proton oscillator-detector similar to that of N. J. Hopkins, Rev. Sci. Instr. **20**, 401 (1949), was used with a water-filled coil.

TABLE II. Comparison of theory and experiment in Λ doubling of the $N^{14}O$ spectrum.

Line	Experimental separation of Λ components (gauss)	Calculated separation of Λ components (gauss)	Deviation (percent)
H_1	1.48	1.482	+0.13
H_2	1.34	1.352	+0.89
H_3	1.32	1.311	−0.69
H_4	1.54	1.533	−0.46
H_5	1.86	1.858	−0.11
H_6	1.49	1.492	+0.13
H_7	1.82	1.829	+0.49
H_8	1.97	1.967	−0.15
H_9	2.01	2.009	−0.05

is estimated as 2 percent of the lambda-component separations. Certainly, the main features of the variation of the doubling with M_J and M_I are as predicted by Karnaugh's theory.

The Molecular Zeeman Effect

In the earlier work[1,2] the molecular Zeeman splittings were a principal point of comparison of theory and experiment. The agreement at that time was quite satisfactory. It has not been improved by using the present data.

In the earlier work the molecular splittings were expressed as the resonance field values H_I, H_{II}, H_{III} of a hypothetical nitric oxide with spinless nitrogen. To obtain these field values directly from experimental data one must express the hyperfine couplings in field units. This method is only approximate and is not used here. Rather, the values of H_2, H_5, H_8 are compared with a new calculation of the molecular energies $E_{ab}(M_J)+\Delta E_{ab}(M_J)$ which proceeds directly from Eq. (1) using the calculation method of reference 2. As before, the value 123.8 cm^{-1} was taken for the $\mathbf{S}\cdot\mathbf{\Lambda}$ coupling constant.

The results of this calculation are given in Table III.

As a check on the method of energy calculation of reference 2, a second calculation of $E_{ab}(M_J)+\Delta E_{ab}(M_J)$ was made from the seventh-order determinant which arises when all seven eigenfunctions of reference 2 are combined at once. For a field of 8425.71 gauss and $M_J=\frac{1}{2}$ the energies calculated by the two methods differed by only 0.012 percent. Apparently then, the approximations in the calculation method are not responsible for the deviation between theory and experiment.

TABLE III. Molecular Zeeman splitting in the $N^{14}O$ spectrum at 9270.65 Mc/sec.

	H_5	(H_5-H_3)	(H_5-H_2)
Theory	8537.50 gauss	103.07 gauss	104.54 gauss
Experiment	8528.28±0.1	101.43±0.05	102.57±0.05
Percent diff.	0.11 percent	1.6 percent	1.9 percent

The molecular coupling constant of $\mathbf{S}\cdot\mathbf{\Lambda}$, which is given from the band spectrum analysis of the doublet splitting, is not known very accurately. We have used 123.8 cm^{-1} for this constant. A value of 124.2 cm^{-1} has been used[9] most recently in analysis of the rotation-vibration spectrum. The constant may be in error by as much as 0.5 cm^{-1}, and it is important to know how this affects the theoretical value for H_5. This has not been determined using the complete theory of reference 2 but only with the stage 1 approximation using Hill's formula.[10] This formula predicts H_5 quite well, although it does not give good values for the H_5-H_3 and H_5-H_2 differences. With Hill's formula an increase of 4.0 cm^{-1} in the coupling constant is required to decrease the theoretical 8537.5 gauss for H_5 resonance by the required 9.2 gauss. Such an increase in the coupling constant is out of the question, and it is concluded that this cannot account for the discrepancy.

A final point should be mentioned in connection with the H_5 discrepancy. In reference 2, the orbital Bohr magneton was used for both orbital and spin contributions to the magnetic energies. A better procedure would include the spin moment anomaly. For pure case (*a*) the *g* factor of a $J=\frac{3}{2}$, $^2\Pi_{\frac{3}{2}}$ level is $(4/5)(1+2\delta/3)$ rather than the usual 4/5; here δ is the spin moment anomaly 0.0011454. Since we are rather near to case (*a*), it is expected that the theoretical resonance fields of Table III will be multiplied by a factor $(1+2\delta/3)^{-1}=0.999237$. This reduces the theoretical H_5 to 8531.0 gauss. Thus, the spin-moment correction together with an increase of about 1 cm^{-1} in the $\mathbf{S}\cdot\mathbf{\Lambda}$ coupling constant would remove the H_5 discrepancy in Table III. It is known, however, that other points in the theory of reference 2 should be reexamined if theoretical precision of 0.1 percent is desired.

Theory of the Hyperfine Structure

The hyperfine structure of the $^2\Pi_{\frac{3}{2}}$ state of $N^{14}O^{16}$ in high magnetic fields has been developed in two previous papers.[2,3] The present calculation was performed in an attempt to provide a more stringent test of the theory of reference 3 by using the data which resulted from the lambda-doubling measurements. The lambda doubling itself does not appear in these calculations; it is assumed that it splits each hyperfine level symmetrically and that the hyperfine constants can be evaluated from the nine-line spectrum consisting of the arithmetic mean positions of the nine lambda doublets.

It is convenient to represent the energy of one of these nine components as $W(M_J,M_I)$ such that

$$W(M_J,M_I)=E_{ab}(M_J)+\Delta E_{ab}(M_J)+E_I(M_J,M_I)+\Delta E_I(M_J,M_I)+E_q(M_J,M_I). \quad (1)$$

Here $E_{ab}(M_J)$ and $\Delta E_{ab}(M_J)$ are the first- and second-order molecular Zeeman energies; they are

[9] R. H. Gillette and E. H. Eyster, Phys. Rev. **56**, 1113 (1939).
[10] E. L. Hill, Phys. Rev. **34**, 1507 (1929).

functions of the magnetic field strength and of M_J, the quantum number representing the component of the total angular momentum of the molecule (exclusive of nuclear spin) in the direction of the magnetic field. Expressions for $E_{ab}(M_J)$ and $\Delta E_{ab}(M_J)$ as power series in M_J and the magnetic field strength are given in Sec. 3 of reference 2.

The term $E_q(M_J,M_I)$ represents the energy due to the quadrupole moment of the N^{14} nucleus, M_I being the quantum number associated with the component of nuclear spin of N^{14} in the direction of the magnetic field. In the approximation employed here, $E_q(M_J,M_I)$ is independent of the magnetic field strength. It is given by Eq. 14 of reference 2 which, for the $J=\frac{3}{2}$, $^2\Pi_{\frac{1}{2}}$ level, is

$$E_q(M_J,M_I)=\frac{|e|Qq}{60}\left(\frac{15}{4}-3M_J^2\right)(2-3M_I^2), \quad (2)$$

where $|e|Qq$ is the nuclear quadrupole coupling constant defined in the manner of Bardeen and Townes.[11]

The quantities $E_I(M_J,M_I)$ and $\Delta E_I(M_J,M_I)$ are the first- and second-order interactions between the magnetic moment of the N^{14} nucleus and the magnetic fields due to the spin and the orbital motion of the $^2\pi$ electron. Application of standard perturbation theory yields the usual expressions

$$E_I(M_J,M_I)=\langle\Psi(M_J,M_I)|\mathcal{H}_I|\Psi(M_J,M_I),\rangle \quad (3)$$

$$\Delta E_I(M_J,M_I)=\frac{\langle\Psi(M_J,M_I)|\mathcal{H}_I|\Psi(M_J+1,M_I-1)\rangle^2}{W(M_J,M_I)-W(M_J+1,M_I-1)}+\frac{\langle\Psi(M_J,M_I)|\mathcal{H}_I|\Psi(M_J-1,M_I+1)\rangle^2}{W(M_J,M_I)-W(M_J-1,M_I+1)}, \quad (4)$$

in which $\Psi(M_J,M_I)$ is the state function of $^2\Pi_{\frac{1}{2}}$ level in the absence of any nuclear interaction, and $\mathcal{H}_I$ is the operator representing the interaction.

The state function, $\Psi(M_J,M_I)$, can be expressed as a product of an electronic, a vibrational, a rotational, and a nuclear part. We shall write the product of the first three of these as $\Psi(M_J)$ and the nuclear part as $|I,M_I\rangle$, I being the quantum number associated with the nuclear spin. Moreover, since we are here concerned with only one electronic and vibrational level, we may write $\Psi(M_J)$ as a linear combination of a basic set of eigenfunctions each having the same electronic and vibrational parts but different rotational parts. So that use could be made of matrix elements already computed by Hill,[10] the rotational parts were taken in reference 2 to be functions representing Hund's case (b), in which the rotational quantum number K, the total electron spin S, the total angular momentum quantum number J, and its component M_J, in the direction of the magnetic field are sharp. These eigenfunctions are written $|K,J,M_J\rangle$; it was found that a combination of four of them yield a sufficiently accurate expression for $\Psi(M_J)$. This combination is (see reference 3, Eq. 5)

$$\Psi(M_J)=a|1,\tfrac{3}{2},M_J\rangle+b|2,\tfrac{3}{2},M_J\rangle+c|2,5/2,M_J\rangle+d|3,5/2,M_J\rangle, \quad (5)$$

where the coefficients are functions of M_J and of the magnetic field strength; their values are such that $\Psi(M_J)$, although compounded of "case b" eigenfunctions, is fairly close to an eigenfunction representing Hund's case (a). Thus the state functions to be used in (3) and (4) have the form

$$\Psi(M_J,M_I)=a|1,\tfrac{3}{2},I,M_J,M_I\rangle+b|2,\tfrac{3}{2},I,M_J,M_I\rangle+c|2,5/2,I,M_J,M_I\rangle+d|3,5/2,I,M_J,M_I\rangle, \quad (6)$$

in which the product $|K,J,M_J\rangle|I,M_I\rangle$ is written $|K,J,I,M_J,M_I\rangle$, explicit dependence on electron spin, vibrational, and electronic quantum numbers being supressed.

The interaction operator, averaged over electronic and vibrational coordinates is given by [reference 3, Eq. (9) with an obvious change in notation]

$$\mathcal{H}_I=A\mathbf{I}\cdot\mathbf{\Lambda}+B[S_zI_z(3\cos^2\theta-1)+\tfrac{3}{2}I_z\sin\theta\cos\theta(iS^+e^{-i\psi}-iS^-e^{i\psi})+\tfrac{3}{2}S_z\sin\theta\cos\theta(iI^+e^{-i\psi}-iI^-e^{i\psi})-\tfrac{3}{4}\sin^2\theta(S^+I^+e^{-2i\psi}+S^-I^-e^{2i\psi})-\tfrac{1}{4}(3\cos^2\theta-1)(S^+I^-+S^-I^+)], \quad (7)$$

where $\mathbf{I}$ is the nuclear spin operator, $\mathbf{\Lambda}$ the operator representing the component of orbital angular momentum along the internuclear axis of the molecule, $\mathbf{S}$ the electronic spin operator; the Euler angles θ and ψ specify the orientation of the molecule with respect to the direction of the magnetic field, and the constants A and B are averages over electronic and vibrational coordinates. Specifically,

$$A\equiv\langle 4\mu_0\mu_I/r^3\rangle_{\text{elec, vib}};\qquad B\equiv\langle 2\mu_0\mu_I(3\cos^2\chi-1)/r^3\rangle_{\text{elec, vib}}, \quad (8)$$

where μ_0 is the Bohr magneton; μ_I is one-half the magnitude of the nuclear magnetic moment of N^{14}, r is the distance between the $^2\pi$ electron and the N^{14} nucleus, and χ is the third Euler angle of the set θ,ψ,χ.

In a recent paper[12] Frosch and Foley obtained a relativistic expression for the hyperfine interaction operator; this expression reduces to (7) in the nonrelativistic approximation. In the absence of any information about the electronic part of the state function for the $^2\Pi_{\frac{1}{2}}$ state of NO, it is difficult to determine whether or not a relativistic expression is needed. In this particular case, provided only effects diagonal in Λ are considered, the sole difference between the

[11] J. Bardeen and C. H. Townes, Phys. Rev. 73, 97 (1948).

[12] R. A. Frosch and H. M. Foley, Phys. Rev. 88, 1337 (1952).

relativistic and nonrelativistic expressions lies in the interpretation of the constants A and B. Application of Eqs. (1) of reference 3 allows (7) to be written in a form appropriate for case (a) eigenfunctions:

$$\begin{aligned}\mathcal{H}_I &= A\mathbf{I}\cdot\mathbf{\Lambda}+B(-\mathbf{I}\cdot\mathbf{S}+3I_{Z'}S_{Z'})\\ &= (A\Lambda+2B\Sigma)\mathbf{I}\cdot\mathbf{k}'-B(I_{X'}S_{X'}+I_{Y'}S_{Y'}),\end{aligned}\quad (7')$$

where the components of $\mathbf{I}$ and $\mathbf{S}$ are now expressed in a coordinate system $X'Y'Z'$ fixed in the molecule, Z' being the direction of the intermolecular axis and $\mathbf{k}'$ being a unit vector in this direction. The expression (7′) is identical in form with Frosch and Foley's Eq. (10.4), the two differing only in the definition of the coefficients of the operators. The relativistic theory shows that $(A\Lambda+2B\Sigma)$ should be replaced by

$$[a\Lambda+(b+c)\Sigma],$$

and B by $-b$, where

$$a\equiv A;\quad b\equiv -B+\left\langle\frac{8\mu_0\mu_I\delta(r)}{3r^3}\right\rangle_{\text{elec, vib}};\quad c\equiv 3B.$$

Since only the two coefficients can be obtained from analysis of the data, a test of whether or not there is any spherically symmetric component to the electronic part of the state function ($3b+c\neq 0$) appears impossible from the first-order hyperfine interaction.

Matrix elements of the first term of (7) between eigenfunctions $|K_1, J_1, M_J\rangle$ and $|K_2, J_2, M_J\rangle$ are given by Hill;[10] those of the remaining terms (except the one between $|K, J, M_J\rangle$ and $|K+2, J+1, M_J\rangle$ appear in reference 3). However, it was discovered during the analysis of the present data that the phases of the eigenfunctions employed in calculating the latter elements are inconsistent with those used (implicitly) by Hill. As a result, the second and third of Eqs. (11), reference 3, should be multiplied by (-1) if they are to be used in conjunction with the matrix elements given by Hill.

An exact calculation of (3) and (4) requires use of the following additional matrix elements:

$$\langle K, J, \tfrac{1}{2}, I, M_J, M_I|\mathcal{H}_I|K+2, J+1, \tfrac{1}{2}, I, M_J, M_I\rangle = \frac{3M_I}{2(K+1)(K+2)(2K+3)}\{[(K+1)^2-\Lambda^2][(K+2)^2-\Lambda^2][(J+1)^2-M_J^2]\}^{\frac{1}{2}},\quad (9)$$

$$\langle K, J, \tfrac{1}{2}, I, M_J, M_I|\mathcal{H}_I|K, J, \tfrac{1}{2}, I, M_J\pm 1, M_I\mp 1\rangle = [(1\pm M_I)(2\mp M_I)(J\mp M_J)(J\pm M_J+1)]^{\frac{1}{2}} \times\left\{\frac{A\Lambda^2}{(2K+1)(J+\frac{1}{2})}+\frac{B}{(4K^2-1)(2K+3)}\left[\frac{3\Lambda^2}{K(K+1)}-1\right]\left[-\tfrac{3}{2}(2K+1)+(J-K)(4K^2+K+3)\right]\right\},\quad (10)$$

$$\langle K, J, \tfrac{1}{2}, I, M_J, M_I|\mathcal{H}_I|K+1, J, \tfrac{1}{2}, I, M_J\pm 1, M_I\mp 1\rangle = \frac{A+3B\Lambda}{(K+1)(2K+1)(2K+3)}\{(1\pm M_I)(2\mp M_I)(J\mp M_J)(J\pm M_J+1)[(K+1)^2-\Lambda^2]\}^{\frac{1}{2}},\quad (11)$$

$$\langle K, J, \tfrac{1}{2}, I, M_J, M_I|\mathcal{H}_I|K, J+1, \tfrac{1}{2}, I, M_J\pm 1, M_I\mp 1\rangle = \pm\frac{2A-(K^2+K-3)B}{4K(K+1)(2K+1)}\{(1\pm M_I)(2\mp M_I)(J\pm M_J+1)(J\pm M_J+2)\}^{\frac{1}{2}},\quad (12)$$

$$\langle K, J, \tfrac{1}{2}, I, M_J, M_I|\mathcal{H}_I|K+1, J+1, \tfrac{1}{2}, I, M_J\pm 1, M_I\mp 1\rangle = \pm\{(1\pm M_I)(2\mp M_I)(J\pm M_J+1)(J\pm M_J+2)[(K+1)^2-\Lambda^2]\}^{\frac{1}{2}} \times\left\{-\frac{A}{4(J+1)(K+1)}+\frac{\Lambda B}{K(K+1)(K+2)(2K+1)(2K+3)}\left[-3(J-K)(K^2+2K+\tfrac{3}{4})+\frac{9}{4}(K+1)\right]\right\},\quad (13)$$

$$\langle K, J, \tfrac{1}{2}, I, M_J, M_I|\mathcal{H}_I|K+2, J+1, \tfrac{1}{2}, I, M_J\pm 1, M_I\mp 1\rangle = \mp\frac{3B}{4(K+1)(K+2)(2K+3)}\{[(K+1)^2-\Lambda^2][(K+2)^2-\Lambda^2](J\pm M_J+1)(J\pm M_J+2)\}^{\frac{1}{2}}.\quad (14)$$

The foregoing have been computed by applying the angular momentum addition laws given by Condon and Shortley[13] to the matrix elements calculated by Rademacher and Reiche.[14] Phases have been chosen so as to be consistent with the elements computed by Hill.

The Hyperfine Constants

If the theory being applied is correct, it should be possible to find a pair of constants A and B [Eqs. (7) and (8)] such that the three following difference equations are satisfied for $W(M_J,M_I)$:

$$\begin{aligned}
&[W(\tfrac{3}{2},1)-W(\tfrac{1}{2},1)]_{H_1}\\
&\quad-[W(\tfrac{3}{2},-1)-W(\tfrac{1}{2},-1)]_{H_3}=\nu_1-\nu_3,\\
&[W(\tfrac{1}{2},1)-W(-\tfrac{1}{2},1)]_{H_4}\\
&\quad-[W(\tfrac{1}{2},-1)-W(-\tfrac{1}{2},-1)]_{H_6}=\nu_4-\nu_6,\\
&[W(-\tfrac{1}{2},1)-W(-\tfrac{3}{2},1)]_{H_7}\\
&\quad-[W(-\tfrac{1}{2},-1)-W(-\tfrac{3}{2},-1)]_{H_9}=\nu_7-\nu_9.
\end{aligned}\tag{15}$$

The $W(M_J,M_I)$ are given by (1), the subscripts H_i refer to the nine line positions of Table I, and the ν_i are the microwave frequencies at which each of the measurements were made.

Unfortunately, the best set of values of A and B obtained from (15) are inconsistent with the definitions (8) which require $A>0$; $-(1/2)A<B\leqslant A$. This difficulty may arise from the fact that the value of B is very sensitive to experimental errors in magnetic field strength differences H_1-H_3, H_4-H_6 and H_7-H_9 (a variation of 1 percent in these differences can change the value of B by a factor of several hundred), or it may be due to some small perturbation which we have failed to include. The fact that seemingly reasonable values of A and B were obtained from the earlier analysis[3] is apparently an accident, due to the inconsistency in phases mentioned above.

[13] E. U. Condon and G. H. Shortley, *Theory of Atomic Spectra* (Cambridge University Press, London, 1951), second edition.
[14] H. Rademacher and F. Reiche, Z. Physik 41, 453 (1927).

If, as is suggested by (7′), we recast (15) in the form of three equations in $(A+B)$ and B, we find that the value of $(A+B)$ is relatively insensitive to the field strength errors and small perturbations which may have been neglected. In fact, $(A+B)$ changes by less than 3.7 percent, as B varies over the whole range of values consistent with the definitions (8).

In view of these considerations, we calculate only the sum $A+B$ which is fairly insensitive to limitations in theory and experiment. In the calculation B is assumed negligible as compared with $A+B$. Frosch and Foley suggested this assumption from examination of (7′). Use of (15) yields three independent values of $A+B$. None differ from $A+B=0.00247$ cm$^{-1}=74.1$ Mc/sec by more than 0.9 percent. Frosch and Foley obtained $[a+\frac{1}{2}(b+c)]=0.002607$ cm^{-1} using the former data[1] and a basic set of eigenfunctions corresponding to case (*a*).

Two values of the quadrupole interaction constant were determined from the equations

$$\begin{aligned}
&[W(\tfrac{3}{2},1)-W(\tfrac{1}{2},1)]_{H_1}\\
&\quad+[W(\tfrac{3}{2},-1)-W(\tfrac{1}{2},-1)]_{H_3}\\
&\quad-2[W(\tfrac{3}{2},0)-W(\tfrac{1}{2},0)]_{H_2}=\nu_1+\nu_3-2\nu_2,\\
&[W(-\tfrac{1}{2},1)-W(-\tfrac{3}{2},1)]_{H_7}\\
&\quad+[W(-\tfrac{1}{2},-1)-W(-\tfrac{3}{2},-1)]_{H_9}\\
&\quad-2[W(-\tfrac{1}{2},0)-W(-\tfrac{3}{2},0)]_{H_8}=\nu_7+\nu_9-2\nu_8.
\end{aligned}\tag{16}$$

The values are $|e|Qq=-0.000076$ cm$^{-1}=-2.28$ Mc/sec and $|e|Qq=-0.000073$ cm$^{-1}=-2.19$ Mc/sec. The mean value $|e|Qq=-0.000075$ cm$^{-1}=-2.23$ Mc/sec agrees with both determinations to within the limits of error of the experiment which is estimated as ± 0.06 Mc/sec in the value of $|e|Qq$. These results are relatively insensitive to the values of A and B. Hence, provided the assumption that the lambda doubling splits each hyperfine line symmetrically is not grossly in error, the sign of the quadrupole coupling constant $|e|Qq$ is definitely negative.

Magnetic Properties of the Hydrogen Molecules

EIICHI ISHIGURO, *Physics Department, Ochanomizu University, Tokyo, Japan*

AND

SHOICHIRO KOIDE, *Institute of Physics, College of General Education, University of Tokyo, Tokyo, Japan*

(Received July 6, 1953)

The diamagnetic susceptibility χ, the rotational magnetic moment μ_r, the magnetic shielding constant σ, and the spin-rotational magnetic interaction constant H_r of the hydrogen molecule are calculated theoretically. The energy of the rotating molecule, under an external magnetic field, is calculated by means of the variational method using 7+2 term wave function. By expanding the energy in powers of the appropriate parameters, χ, μ_r, σ, and H_r are determined and the relations between these quantities are obtained. Comparison with Ramsey's experiments is also made.

1. INTRODUCTION

As a result of the increasing accuracy of microwave and molecular beam measurements, it is found that the apparent magnetic moment of a nucleus has different values when it is contained in different molecules. This phenomenon, called "chemical shift," arises from the fact that the surrounding electric charges create a magnetic field at the position of the nucleus in addition to the external one. Based on Wick's theory,[1] Ramsey has treated this problem in a series of papers.[2]

It seems to be important to study the chemical shift for the hydrogen molecule, since accurate experimental data are available for this molecule, and its nuclear magnetic moment is used as the standard for the measurement of magnetic field strength. The calculations were made by Ramsey[3] and by Hylleraas and Skavlem.[3] In all these calculations, however, the perturbation method was adopted, and the difficulty lies chiefly in the fact that the second-order terms cannot be evaluated accurately.

On the other hand, one of the present authors has succeeded in calculating the polarizability of the hydrogen molecule and in overcoming the difficulty by using the variational method.[4] The aim of the present paper is to show that the same method is also applicable in principle to the calculation of the magnetic properties of this molecule. We have calculated the chemical shift σ (magnetic shielding constant) as well as the other magnetic quantities, i.e., the diamagnetic susceptibility χ, the rotational magnetic moment μ_r, and the spin-rotational magnetic interaction constant H_r, and found that the well-known relations between these quantities can be formulated quite fundamentally.

For the purpose of examining the usefulness of this method, a preliminary calculation has been made and rather good results have been obtained.

2. HAMILTONIAN

The Hamiltonian of a hydrogen molecule placed in the magnetic field $\mathbf{H}$ is given by

$$\mathcal{H}_{\text{tot}}=\frac{1}{2M}\sum_{a,b}\left(\mathbf{P}_k-\frac{e}{c}\mathbf{A}_k\right)^2+\frac{1}{2m}\sum_{j=1}^{2}\left(\mathbf{p}_j+\frac{e}{c}\mathbf{A}_j\right)^2 +V-\mu'\mathbf{H}(\boldsymbol{\sigma}_a+\boldsymbol{\sigma}_b)+f(\mu'\boldsymbol{\sigma}_a,\mu'\boldsymbol{\sigma}_b), \quad (1)$$

where we neglect the electronic spin which has no influence on our final result within the accuracy of our approximations. $\boldsymbol{\sigma}_a$ and $\boldsymbol{\sigma}_b$ denote the Pauli matrices of the two protons a and b, respectively, and μ' is the magnitude of the magnetic moment of the proton. We may adopt the representation in which the total nuclear spin of the molecule can be specified by a set of quantum numbers (I,m_I), and take only the diagonal elements. Then the x and y components of the $\boldsymbol{\sigma}$'s may be omitted. For the states with $m_I=1, 0, -1$, the expectation values of σ_{az} and σ_{bz} are $\langle\sigma_{az}\rangle=\langle\sigma_{bz}\rangle=1, 0, -1$, respectively. Therefore,

$$\mu'\sigma_{az}=\mu'\sigma_{bz}\equiv\mu. \quad (2)$$

The first term of $\mathcal{H}_{\text{tot}}$ represents the kinetic energy of the two-proton system and, neglecting the term for

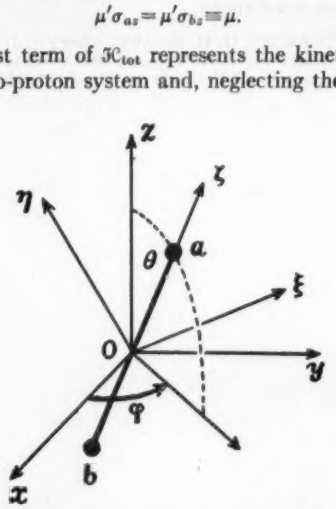

FIG. 1. Relation between the fixed and moving coordinates. a and b denote the nuclei. The ξ axis is taken to lie in the xy plane.

[1] G. C. Wick, Phys. Rev. **73**, 51 (1948); Z. Physik **85**, 25 (1933).
[2] N. F. Ramsey, Phys. Rev. **77**, 567 (1950); **78**, 699 (1950); **83**, 540 (1951); **85**, 60 (1952); 243 (1952).
[3] E. Hylleraas and S. Skavlem, Phys. Rev. **79**, 117 (1950).
[4] Ishiguro, Arai, Mizushima, and Kotani, Proc. Phys. Soc. (London) **A65**, 178 (1952).

the motion of the center of gravity, we can write in terms of polar coordinates:

$$\mathcal{H}_N=-\frac{\hbar^2}{MR^2}\left\{\frac{\partial}{\partial R}\left(R^2\frac{\partial}{\partial R}\right)+\cot\theta\frac{\partial}{\partial\theta}+\frac{\partial^2}{\partial\theta^2}+\csc^2\theta\frac{\partial^2}{\partial\varphi^2}\right\}$$
$$-\frac{e\hbar}{2MC}H\left(-i\frac{\partial}{\partial\varphi}\right)-\frac{e\hbar}{2MC}\frac{2}{R^2}2\mu\left(-i\frac{\partial}{\partial\varphi}\right), \quad (3)$$

where R denotes the internuclear distance of this molecule. In this derivation, the z direction is chosen to be parallel to the external magnetic field, the vector potentials $\mathbf{A}_a$ and $\mathbf{A}_b$ are assumed to be expressed in terms of the coordinates with their origin at the center of the gravity as

$$A_{aX}=-\tfrac{1}{2}HY_a-\frac{2\mu}{R^3}Y_a, \qquad A_{bX}=-\tfrac{1}{2}HY_b-\frac{2\mu}{R^3}Y_b,$$

$$A_{aY}=\tfrac{1}{2}HX_a+\frac{2\mu}{R^3}X_a, \quad \text{and} \quad A_{bY}=\tfrac{1}{2}HX_b+\frac{2\mu}{R^3}X_b, \quad (4)$$

$$A_{aZ}=0, \qquad A_{bZ}=0,$$

and the small terms proportional to the square of A's are neglected. The last terms of Eqs. (4) are the components of the vector potential due to the presence of the proton magnetic moment.

The second term of $\mathcal{H}_{tot}$ is the kinetic energy of the two electrons and, together with the third term V, the electrostatic potential energy of the four particles, can be expressed most readily in terms of the ξ, η, ζ reference system which has the ζ axis parallel to the axis of molecule and travels with the nuclei (see Fig. 1). We may suppose the ξ axis to lie in the xy plane without loss of generality. Then the moving and fixed systems are connected by

$$\begin{aligned}\xi&=-x\sin\varphi+y\cos\varphi,\\ \eta&=-x\cos\theta\cos\varphi-y\cos\theta\sin\varphi+z\sin\theta,\\ \zeta&=x\sin\theta\cos\varphi+y\sin\theta\sin\varphi+z\cos\theta.\end{aligned} \quad (5)$$

Using these coordinates and adopting as the vector potentials A_1, A_2, the following expressions:

$$\begin{aligned}A_{jx}&=-(1/2)Hy_j-(\mu/r_{aj}^3)y_{aj}-(\mu/r_{bj}^3)y_{bj}\\ A_{jy}&=(1/2)Hx_j+(\mu/r_{aj}^3)x_{aj}+(\mu/r_{bj}^3)x_{bj}\\ A_{jz}&=0,\end{aligned} \quad (6)$$

where

$$x_{aj}=x_j-X_a, \text{ etc.}, \quad (6a)$$

we obtain

$$\mathcal{H}_e=\frac{1}{2m}\sum_j\left(\mathbf{p}_j+\frac{e}{c}\mathbf{A}_j\right)^2+V \quad (7)$$
$$=\mathcal{H}^{(0)}+H\mathcal{H}^{(1)}+\mu\mathcal{H}^{(2)}+H^2\mathcal{H}^{(4)}+H\mu\mathcal{H}^{(5)}+\cdots,$$

where

$$\mathcal{H}^{(0)}=-\frac{\hbar^2}{2m}\sum_j\left(\frac{\partial^2}{\partial\xi_j^2}+\frac{\partial^2}{\partial\eta_j^2}+\frac{\partial^2}{\partial\zeta_j^2}\right)+V, \quad (8a)$$

$$\mathcal{H}^{(1)}=\frac{e\hbar}{2mci}\sum_j\left\{\sin\theta\left(\zeta_j\frac{\partial}{\partial\xi_j}-\xi_j\frac{\partial}{\partial\zeta_j}\right)\right.$$
$$\left.+\cos\theta\left(\xi_j\frac{\partial}{\partial\eta_j}-\eta_j\frac{\partial}{\partial\xi_j}\right)\right\}, \quad (8b)$$

$$\mathcal{H}^{(2)}=\frac{e\hbar}{2mci}\sum_j\left[\left\{2\sin\theta\left(\zeta_j\frac{\partial}{\partial\xi_j}-\xi_j\frac{\partial}{\partial\zeta_j}\right)\right.\right.$$
$$\left.\left.+2\cos\theta\left(\xi_j\frac{\partial}{\partial\eta_j}-\eta_j\frac{\partial}{\partial\xi_j}\right)\right\}\left(\frac{1}{r_{aj}^3}+\frac{1}{r_{bj}^3}\right)\right.$$
$$\left.-R\sin\theta\left(\frac{1}{r_{aj}^3}-\frac{1}{r_{bj}^3}\right)\frac{\partial}{\partial\xi_j}\right], \quad (8c)$$

$$\mathcal{H}^{(4)}=\frac{e^2}{8mc^2}\sum_j\{\xi_j^2+(\eta_j\cos\theta-\zeta_j\sin\theta)^2\}, \quad (8d)$$

$$\mathcal{H}^{(5)}=\frac{e^2}{2mc^2}\sum_j\left[\{\xi_j^2+(\eta_j\cos\theta-\zeta_j\sin\theta)^2\}\left(\frac{1}{r_{aj}^3}+\frac{1}{r_{bj}^3}\right)\right.$$
$$\left.+\frac{R}{2}\sin\theta(\zeta_j\sin\theta-\eta_j\cos\theta)\left(\frac{1}{r_{aj}^3}-\frac{1}{r_{bj}^3}\right)\right]. \quad (8e)$$

In terms of the same coordinate system, we may also take the wave function as the product of two parts:

$$\psi=\Phi(R,\theta,\varphi)\Psi(\xi_1,\eta_1,\zeta_1,\xi_2,\eta_2,\zeta_2;R,\theta). \quad (9)$$

Then, since ξ, η, ζ depend upon θ and φ through the relations (5), the differential operators appearing in the Eq. (3) have to be transformed as

$$\frac{\partial}{\partial\theta}\rightarrow\frac{\partial}{\partial\theta}+\sum_j\left(\frac{\partial\xi_j}{\partial\theta}\frac{\partial}{\partial\xi_j}+\frac{\partial\eta_j}{\partial\theta}\frac{\partial}{\partial\eta_j}+\frac{\partial\zeta_j}{\partial\theta}\frac{\partial}{\partial\zeta_j}\right)$$
$$=\frac{\partial}{\partial\theta}-\sum_j\left(\eta_j\frac{\partial}{\partial\zeta_j}-\zeta_j\frac{\partial}{\partial\eta_j}\right), \quad (10a)$$

$$\frac{\partial}{\partial\varphi}\rightarrow\frac{\partial}{\partial\varphi}+\sum_j\left(\frac{\partial\xi_j}{\partial\varphi}\frac{\partial}{\partial\xi_j}+\frac{\partial\eta_j}{\partial\varphi}\frac{\partial}{\partial\eta_j}+\frac{\partial\zeta_j}{\partial\varphi}\frac{\partial}{\partial\zeta_j}\right)$$
$$=\frac{\partial}{\partial\varphi}-\sum_j\left\{\sin\theta\left(\zeta_j\frac{\partial}{\partial\xi_j}-\xi_j\frac{\partial}{\partial\zeta_j}\right)\right.$$
$$\left.+\cos\theta\left(\xi_j\frac{\partial}{\partial\eta_j}-\eta_j\frac{\partial}{\partial\xi_j}\right)\right\}. \quad (10b)$$

The physical significance of this transformation is most easily seen if, for instance, we consider the

meaning of the operator $\partial/\partial\varphi$. As is well known, $(-i\partial/\partial\varphi)_{R,\theta,\xi_1,\cdots,\zeta_2}$ represents the z component of the angular momentum of the free rotator composed of two protons. On the other hand, since the new $\partial/\partial\varphi$ means differentiation with R, θ, ξ_1, $\cdots$, and ζ_2 kept constant, the new $-i\partial/\partial\varphi$ is the operator which represents the angular momentum of the two-proton system carrying the two electrons with it. Therefore, the difference between the new and the old $-i\partial/\partial\varphi$ should be equal to the z component of the angular momentum of the two electrons relative to the nuclear system. Equations (10b) express this relation. A similar situation is valid for $\partial/\partial\theta$ *mutatis mutandis*.

Substituting Eqs. (10a) and (10b) into (3) and considering Eq. (1) and $-\mu' H(\sigma_{a_z}+\sigma_{b_z})=-2\mu H$, we finally obtain the total Hamiltonian:

$$\begin{aligned}\mathfrak{H}_{\text{tot}}=&-\frac{\hbar^2}{MR^2}\left\{\frac{\partial}{\partial R}\left(R^2\frac{\partial}{\partial R}\right)+\cot\theta\left(\frac{\partial}{\partial\theta}-iL_\xi\right)\right.\\&\left.+\left(\frac{\partial}{\partial\theta}-iL_\xi\right)^2+\csc^2\theta\left(\frac{\partial}{\partial\varphi}-iL_z\right)^2\right\}\\&+\frac{ie\hbar}{2Mc}H\left(\frac{\partial}{\partial\varphi}-iL_z\right)+\frac{ie\hbar}{2Mc}\frac{4\mu}{R^3}\left(\frac{\partial}{\partial\varphi}-iL_z\right)\\&-2\mu H+\mathfrak{H}^{(0)}+H\mathfrak{H}^{(1)}+\mu\mathfrak{H}^{(2)}\\&+H^2\mathfrak{H}^{(4)}+H\mu\mathfrak{H}^{(5)},\end{aligned}\tag{11}$$

where use has been made of the following abbreviations:

$$\begin{aligned}L_\xi&=l_\xi{}^1+l_\xi{}^2=-i\sum_j\left(\eta_j\frac{\partial}{\partial\zeta_j}-\zeta_j\frac{\partial}{\partial\eta_j}\right),\\L_\eta&=l_\eta{}^1+l_\eta{}^2=-i\sum_j\left(\zeta_j\frac{\partial}{\partial\xi_j}-\xi_j\frac{\partial}{\partial\zeta_j}\right),\\L_\zeta&=l_\zeta{}^1+l_\zeta{}^2=-i\sum_j\left(\xi_j\frac{\partial}{\partial\eta_j}-\eta_j\frac{\partial}{\partial\xi_j}\right),\\L_z&=L_\eta\sin\theta+L_\zeta\cos\theta.\end{aligned}\tag{12}$$

3. PROCEDURE

The Hamiltonian given by (11) seems to be unmanageable. We will, therefore, treat it in the following manner:

(i) First we determine the internal wave function $\Psi_0(\xi_1,\eta_1,\zeta_1,\xi_2,\eta_2,\zeta_2;R)$ and its lowest eigenvalue E_0 in the case $H=0$ and $\mu=0$, by solving the equation

$$\mathfrak{H}^{(0)}\Psi_0=E_0\Psi_0.\tag{13}$$

As is well known, this calculation has been made by many investigators, including one of the present authors.[5] We can, therefore, utilize the results already obtained.

[5] E. Ishiguro, J. Phys. Soc. Japan 3, 129 (1948).

(ii) Using the above Ψ_0, one should be able to determine the vibration-rotation wave function Φ_0 in the case $H=\mu=0$. A rigorous treatment might be difficult, of course. However, it has been shown that to a good approximation Φ_0 can be written as $f(R)\cdot\chi(\theta,\varphi)$, and the rotational eigenfunction $\chi(\theta,\varphi)$ can be expressed by the spherical harmonic $Y_J{}^{m_J}$ which has the eigenvalue J for the angular momentum and m_J for its z component.

(iii) Finally, introducing H and μ, we calculate the internal wave function $\Psi(\xi_1,\eta_1,\zeta_1,\xi_2,\eta_2,\zeta_2;R,\theta)$ by means of the variation method. In this step, χ is assumed to be not considerably different from the one obtained in (ii). Then we can safely replace the operator $-i\partial/\partial\varphi$ (z component of the angular momentum of nuclear system) with its eigenvalue m_J. From the first term of (11) we pick out the term proportional to m_J and consider it as a perturbation in addition to the terms $H\mathfrak{H}^{(1)}$, $\mu\mathfrak{H}^{(2)}$, $H^2\mathfrak{H}^{(4)}$, and $H\mu\mathfrak{H}^{(5)}$. Then,

$$\begin{aligned}m_J\mathfrak{H}^{(3)}=&-\frac{2\hbar^2m_J}{iMR^2\sin^2\theta}\sum_{j=1}^{2}\left\{\sin\theta\left(\zeta_j\frac{\partial}{\partial\xi_j}-\xi_j\frac{\partial}{\partial\zeta_j}\right)\right.\\&\left.+\cos\theta\left(\xi_j\frac{\partial}{\partial\eta_j}-\eta_j\frac{\partial}{\partial\xi_j}\right)\right\}.\end{aligned}\tag{14}$$

The remaining electronic terms are all neglected because they contribute nothing to the quantities in which we are interested. Thus the Hamiltonian in question becomes

$$\begin{aligned}\mathfrak{H}=&\mathfrak{H}^{(0)}+H\mathfrak{H}^{(1)}+\mu\mathfrak{H}^{(2)}\\&+m_J\mathfrak{H}^{(3)}+H^2\mathfrak{H}^{(4)}+H\mu\mathfrak{H}^{(5)}.\end{aligned}\tag{15}$$

4. APPLICATION OF THE VARIATION METHOD

The electronic wave function of the ground state of the free hydrogen molecule, which is the solution of Eq. (13), has the symmetry ${}^1\Sigma_g{}^+$. When the perturbations are introduced, the electron distribution is deformed and the wave function loses its symmetry. If the perturbations are sufficiently small, the wave function may be expressed as

$$\begin{aligned}\Psi=&\sum_i a_i\theta_i+H\sum_k b_k\varphi_k\\&+\mu\sum_l c_l\psi_l+m_J\sum_n d_n\chi_n+\cdots,\end{aligned}\tag{16}$$

where the θ's have the symmetry ${}^1\Sigma_g{}^+$, and φ_k, ψ_l, and χ_n have the other symmetries. Normalization of Ψ gives

$$\begin{aligned}1=&\sum_{i,j}a_i{}^*a_jS_{ij}+H^2\sum_{k,k'}b_k{}^*b_{k'}S_{kk'}\\&+H\mu\sum_{k,l}(b_k{}^*c_lS_{kl}+c_l{}^*b_kS_{lk})\\&+Hm_J\sum_{k,n}(b_k{}^*d_nS_{kn}+d_n{}^*b_kS_{nk})\\&+\mu m_J\sum_{l,n}(c_l{}^*d_nS_{ln}+d_n{}^*c_lS_{nl})+\cdots,\end{aligned}\tag{17}$$

where $S_{ij}=\int\theta_i^*\theta_j d\tau$, $S_{kl}=\int\varphi_k^*\psi_l d\tau$, etc., and S_{ik}, S_{il}, and S_{in} vanish because of symmetry. The mean value of the Hamiltonian $\mathcal{H}$ with respect to the function Ψ is given by

$$\begin{aligned}E=&\sum a_i^*a_j\mathcal{H}_{ij}^{(0)}+H^2\{\sum a_i^*a_j\mathcal{H}_{ij}^{(4)}+\sum b_k^*b_{k'}\mathcal{H}_{kk'}^{(0)}\\&+\sum(a_i^*b_k\mathcal{H}_{ik}^{(1)}+\text{c.c.})\}\\&+H\mu\{\sum(a_i^*c_l\mathcal{H}_{il}^{(1)}+\text{c.c.})+\sum(a_i^*b_k\mathcal{H}_{ik}^{(2)}+\text{c.c.})\\&+\sum a_i^*a_j\mathcal{H}_{ij}^{(5)}+\sum(b_k^*c_l\mathcal{H}_{kl}^{(0)}+\text{c.c.})\}\\&+Hm_J\{\sum(a_i^*d_n\mathcal{H}_{in}^{(1)}+\text{c.c.})+\sum(a_i^*b_k\mathcal{H}_{ik}^{(3)}\\&+\text{c.c.})+\sum(b_k^*d_n\mathcal{H}_{kn}^{(0)}+\text{c.c.})\}\\&+\mu m_J\{\sum(a_i^*d_n\mathcal{H}_{in}^{(2)}+\text{c.c.})+\sum(a_i^*c_l\mathcal{H}_{il}^{(3)}+\text{c.c.})\\&+\sum(c_l^*d_n\mathcal{H}_{ln}^{(0)}+\text{c.c.})\},\end{aligned}\tag{18}$$

where c.c. means the complex conjugate. We try to find the coefficient of (16) which minimizes E of Eq. (18) under the condition given by (17). Differentiating (18) with respect to a_i^*, b_k^*, c_l^*, and d_n^*, and using the well-known Lagrangian multiplier method, we obtain:

$$\begin{aligned}&\sum_j(\mathcal{H}_{ij}^{(0)}-S_{ij}E)a_j+H^2(\sum_j\mathcal{H}_{ij}^{(4)}a_j+\sum_k\mathcal{H}_{ik}^{(1)}b_k)\\&+H\mu(\sum_l\mathcal{H}_{il}^{(1)}c_l+\sum_k\mathcal{H}_{ik}^{(2)}b_k+\sum_j\mathcal{H}_{ij}^{(5)}a_j)\\&+Hm_J(\sum_n\mathcal{H}_{in}^{(1)}d_n+\sum_k\mathcal{H}_{ik}^{(3)}b_k)\\&+\mu m_J(\sum_n\mathcal{H}_{in}^{(2)}d_n+\sum_l\mathcal{H}_{il}^{(3)}c_l)=0,\end{aligned}\tag{19}$$

$$\begin{aligned}&H^2\{\sum_{k'}(\mathcal{H}_{kk'}^{(0)}-S_{kk'}E)b_{k'}+\sum_i\mathcal{H}_{ki}^{(1)}a_i\}\\&+H\mu\{\sum_i\mathcal{H}_{ki}^{(2)}a_i+\sum_l(\mathcal{H}_{kl}^{(0)}-S_{kl}E)c_l\}\\&+Hm_J\{\sum_n(\mathcal{H}_{kn}^{(0)}-S_{kn}E)d_n+\sum_i\mathcal{H}_{ki}^{(3)}a_i\}=0,\end{aligned}\tag{20}$$

$$\begin{aligned}&H\mu\{\sum_i\mathcal{H}_{li}^{(1)}a_i+\sum_k(\mathcal{H}_{lk}^{(0)}-S_{lk}E)b_k\}\\&+\mu m_J\{\sum_i\mathcal{H}_{li}^{(3)}a_i+\sum_n(\mathcal{H}_{ln}^{(0)}-S_{ln}E)d_n\}=0,\end{aligned}\tag{21}$$

$$\begin{aligned}&\mu m_J\{\sum_i\mathcal{H}_{ni}^{(2)}a_i+\sum_l(\mathcal{H}_{nl}^{(0)}-S_{nl}E)c_l\\&+Hm_J\{\sum_n\mathcal{H}_{ni}^{(1)}a_i+\sum_k(\mathcal{H}_{nk}^{(0)}-S_{nk}E)b_k\}=0.\end{aligned}\tag{22}$$

From these we get at once:

$$\sum_{k'}(\mathcal{H}_{kk'}^{(0)}-E_0S_{kk'})b_{k'}=-\sum_i\mathcal{H}_{ki}^{(1)}a_i^0,\tag{23a}$$

$$\sum_l(\mathcal{H}_{kl}^{(0)}-E_0S_{kl})c_l=-\sum_i\mathcal{H}_{ki}^{(2)}a_i^0,\tag{23b}$$

$$\sum_n(\mathcal{H}_{kn}^{(0)}-E_0S_{kn})d_n=-\sum_i\mathcal{H}_{ki}^{(3)}a_i^0,\tag{23c}$$

$$\sum_k(\mathcal{H}_{lk}^{(0)}-E_0S_{lk})b_k=-\sum_i\mathcal{H}_{li}^{(1)}a_i^0,\tag{23d}$$

$$\sum_n(\mathcal{H}_{ln}^{(0)}-E_0S_{ln})d_n=-\sum_i\mathcal{H}_{li}^{(3)}a_i^0,\tag{23e}$$

$$\sum_l(\mathcal{H}_{nl}^{(0)}-E_0S_{nl})c_l=-\sum_i\mathcal{H}_{ni}^{(2)}a_i^0,\tag{23f}$$

$$\sum_k(\mathcal{H}_{nk}^{(0)}-E_0S_{nk})b_k=-\sum_i\mathcal{H}_{ni}^{(1)}a_i^0.\tag{23g}$$

To satisfy Eqs. (23) consistently, it is necessary that

$$\varphi_k=\psi_i=\chi_n.\tag{24}$$

Thus Eqs. (23) can be reduced to:

$$\sum_{k'}(\mathcal{H}_{kk'}^{(0)}-E_0S_{kk'})b_{k'}=-\sum_i\mathcal{H}_{ki}^{(1)}a_i^0,\tag{25a}$$

$$\sum_l(\mathcal{H}_{kl}^{(0)}-E_0S_{kl})c_l=-\sum_i\mathcal{H}_{ki}^{(2)}a_i^0,\tag{25b}$$

$$\sum_n(\mathcal{H}_{kn}^{(0)}-E_0S_{kn})d_n=-\sum_i\mathcal{H}_{ki}^{(3)}a_i^0.\tag{25c}$$

From Eq. (19) it is seen that a_j can be written as

$$a_j=a_j^0+H^2a_j^{(1)}+H\mu a_j^{(2)}+Hm_Ja_j^{(3)}+\mu m_Ja_j^{(4)}+\cdots.\tag{26}$$

(The proof of this relation is given in the Appendix.) Substituting (26) into (17), we get

$$\begin{aligned}1=&\sum a_i^0a_j^0S_{ij}+H^2[\sum b_k^*b_kS_{kk'}+\sum(a_i^0a_j^{(1)}\\&+a_i^{(1)}a_j^0)S_{ij}]+H\mu[\sum(b_k^*c_lS_{kl}+c_l^*b_kS_{lk})\\&+\sum(a_i^0a_j^{(2)}+a_i^{(2)}a_j^0)S_{ij}]\\&+Hm_J[\sum(b_k^*d_nS_{kn}+d_n^*b_kS_{nk})\\&+\sum(a_i^0a_j^{(3)}+a_i^{(3)}a_j^0)S_{ij}]\\&+\mu m_J[\sum(c_l^*d_nS_{ln}+d_n^*c_lS_{nl})\\&+\sum(a_i^0a_j^{(4)}+a_i^{(4)}a_j^0)S_{ij}]\\&+\cdots.\end{aligned}\tag{27}$$

The expressions in the square brackets should all vanish, because $\sum a_i^0a_j^0S_{ij}=\int|\Psi_0|^2d\tau=1$. Making use of these relations together with (23) and $\Sigma(\mathcal{H}_{ij}^{(0)}-S_{ij}E_0)a_j^0=0$, etc., we can simplify (18) as follows:

$$\begin{aligned}E=&E_0+H^2[\sum a_i^0a_j^0\mathcal{H}_{ij}^{(4)}+\sum a_i^0b_k\mathcal{H}_{ik}^{(1)}]\\&+H\mu[\sum a_i^0a_j^0\mathcal{H}_{ij}^{(5)}+\sum a_i^0c_l\mathcal{H}_{il}^{(1)}\\&+\sum a_i^0b_k\mathcal{H}_{ik}^{(2)}]+Hm_J[\sum a_i^0b_k\mathcal{H}_{ik}^{(3)}\\&+\sum a_i^0d_n\mathcal{H}_{in}^{(1)}]+\mu m_J[\sum a_i^0c_l\mathcal{H}_{il}^{(3)}\\&+\sum a_i^0d_n\mathcal{H}_{in}^{(2)}]+\cdots.\end{aligned}\tag{28}$$

Since $L_i\theta_i=0$, the second term of $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$ can be dropped. Then, we can put

$$\mathcal{H}^{(1)\prime}=\mathcal{H}^{(1)}/\sin\theta=-(eMR^2/4mch)\sin\theta\cdot\mathcal{H}^{(3)},\tag{29}$$

and from (25a) and (25c) we obtain

$$-d_n=(4mch/eMR^2)\sin\theta b_k',\tag{30}$$

where

$$b_k'=b_k/\sin\theta.\tag{31}$$

Similarly,

$$c_l'=c_l/\sin\theta.\tag{32}$$

Thus (25a) becomes

$$\sum(\mathcal{H}_{kk'}^{(0)}-E_0S_{kk'})b_{k'}'=-\sum\mathcal{H}_{ki}^{(1)\prime}a_i^0.\tag{33}$$

Furthermore, if we put

$$\mathcal{H}^{(2)}=\sin\theta\cdot\mathcal{H}^{(2)\prime}+\cos\theta\cdot\mathcal{H}^{(2)\prime\prime},\tag{34}$$

it is shown from the symmetry properties that $\mathcal{H}_{ki}^{(2)\prime\prime}$ vanishes for those values of k to which the non-vanishing $\mathcal{H}_{ki}^{(1)\prime}$ corresponds. On the other hand, if $\mathcal{H}_{ki}^{(1)\prime}=0$, then (33) gives $b_k'=0$ which makes $a_i^0b_k\mathcal{H}_{ki}^{(2)}$ vanish. Therefore, $\mathcal{H}_{ki}^{(2)\prime\prime}$ is of no use for our purposes. Thus (25b) can be written as

$$\sum(\mathcal{H}_{kl}^{(0)}-E_0S_{kl})c_l'=-\sum\mathcal{H}_{ki}^{(2)\prime}a_i^0.\tag{35}$$

It follows from (33) and (35) that

$$\sum a_i^0b_k'\mathcal{H}_{ik}^{(2)\prime}=\sum a_i^0c_l'\mathcal{H}_{il}^{(1)\prime}.$$

Taking account of the above relations, we finally get

$$E=E_0+H^2[\sum a_i^0 a_j^0 \mathcal{H}_{ij}^{(4)}+\sin^2\theta \sum a_i^0 b_k' \mathcal{H}_{ik}^{(1)'}]$$
$$+H\mu[\sum a_i^0 a_j^0 \mathcal{H}_{ij}^{(5)}+\sin^2\theta(2\sum a_i^0 b_k' \mathcal{H}_{ik}^{(2)'}]$$
$$-2Hm_J(4mc\hbar/eMR^2)\sum a_i^0 b_k' \mathcal{H}_{ik}^{(1)'}$$
$$-\mu m_J(4mc\hbar/eMR^2)2\sum a_i^0 b_k' \mathcal{H}_{ik}^{(2)'}+\cdots. \quad (36)$$

On the other hand, the total energy should be written [see Eq. (11)] as:

$$E_{tot}=(\text{Rotational energy})+(\text{vibrational energy})$$
$$-(e\hbar/2Mc)Hm_J-(e\hbar/2Mc)(4/R^3)$$
$$\times\mu m_J-2\mu H+E+\cdots$$
$$=(\text{Rotational energy})+(\text{vibrational energy})$$
$$+E_0+(-\chi/2)H^2-2(1-\sigma)H\mu-Hm_J\mu_r$$
$$-2\mu m_J H_r+\cdots, \quad (37)$$

where χ means the magnetic susceptibility of this molecule, σ the magnetic shielding constant, μ_r the rotational magnetic moment, and H_r the rotational magnetic field (spin-rotational magnetic interaction constant). Comparing (36) with (37), we obtain

$$\chi=-2[\sum a_i^0 a_j^0 \mathcal{H}_{ij}^{(4)}+\sin^2\theta \sum a_i^0 b_k' \mathcal{H}_{ik}^{(1)'}], \quad (38)$$

$$\sigma=(1/2)\sum a_i^0 a_j^0 \mathcal{H}_{ij}^{(5)}+\sin^2\theta \sum a_i^0 b_k' \mathcal{H}_{ik}^{(2)'}, \quad (39)$$

$$\mu_r=(e\hbar/2Mc)+(8mc\hbar/eMR^2)\sum a_i^0 b_k' \mathcal{H}_{ik}^{(1)'}, \quad (40)$$

$$H_r=(e\hbar/2Mc)(2/R^3)$$
$$+(4mc\hbar/eMR^2)\sum a_i^0 b_k' \mathcal{H}_{ik}^{(2)'}. \quad (41)$$

These results depend upon the molecular orientation θ and the internuclear distance R, and the observed values are considered as averages over θ and R.

5. NUMERICAL CALCULATION

The most accurate wave function Ψ_0 of a hydrogen molecule so far obtained is that of James and Coolidge.[6] It is therefore desirable to use it to calculate the χ's and to take as many additional functions as possible. It is, however, very laborious.

TABLE I. Coefficients and energies of the wave functions.

	$R=1.3$	$R=1.4$	$R=1.5$ (atomic units)
a_1^0	2.61810	2.46595	2.29591
a_2^0	0.98603	0.90325	0.82526
a_3^0	−0.70138	−0.67659	−0.64411
a_4^0	−0.49117	−0.59722	−0.66111
a_5^0	0.05030	0.05031	0.05333
a_6^0	0.00090	0.01191	0.01891
a_7^0	−0.04938	−0.03731	−0.02839
E_0 (atomic units)	−1.1553615	−1.1575046	−1.1556648
b_1'	$-0.003335i$	$-0.003257i$	$-0.003144i$
b_2'	$0.001910i$	$0.001819i$	$0.001720i$
c_1'	$0.004416i$	$0.003194i$	$0.002320i$
c_2'	$0.000668i$	$0.000636i$	$0.0005815i$

[6] H. M. James and A. S. Coolidge, J. Chem. Phys. 1, 825 (1933).

TABLE II. The calculated values of χ_{mol}, σ, μ_r, and H_r for $R=1.3$, 1.4, and 1.5 atomic units and for arbitrary θ.[a]

	R (atomic units)		Unit
χ_{mol}	1.3	$-3.4674-0.5974\sin^2\theta+0.1469\sin^4\theta$	10^{-6} erg G^{-2} mol^{-1}
	1.4	$-3.5814-0.7339\sin^2\theta+0.1947\sin^4\theta$	10^{-6} erg G^{-2} mol^{-1}
	1.5	$-3.7206-0.8838\sin^2\theta+0.2485\sin^4\theta$	10^{-6} erg G^{-2} mol^{-1}
σ	1.3	$2.897-0.1515\sin^2\theta+0.0706\sin^4\theta$	10^{-5}
	1.4	$2.774-0.0573\sin^2\theta+0.0683\sin^4\theta$	10^{-5}
	1.5	$2.657+0.0292\sin^2\theta+0.0643\sin^4\theta$	10^{-5}
μ_r	1.3	$1-0.1466$	$e\hbar/2Mc$
	1.4	$1-0.1675$	$e\hbar/2Mc$
	1.5	$1-0.1862$	$e\hbar/2Mc$
H_r	1.3	$31.004+2.138$	Gauss
	1.4	$24.823+1.783$	Gauss
	1.5	$20.182+1.463$	Gauss

[a] The first and the second terms of χ_{mol} come from the first term of (38), and the third terms (high-frequency terms) from the second term of (38). The first two terms of each σ come from the first term of (39), and the last from the second term of (39). The dominant part of H_r (the first terms) is obtained from the first term of (41).

In the present paper, for $\sum a_i^0\theta_i$, we tentatively adopt the seven-term wave function obtained by one of the present authors:[5]

$$\Psi_0=(1/2\pi)\exp\{-\delta(\lambda_1+\lambda_2)\}[2a_1^0+a_2^0(\mu_1^2+\mu_2^2)$$
$$+2a_3^0\mu_1\mu_2+a_4^0(\lambda_1+\lambda_2)+a_5^0(\lambda_1^2+\lambda_2^2)$$
$$+2a_6^0\lambda_1\lambda_2\mu_1\mu_2+2a_7^0\rho\cos(\phi_1-\phi_2)], \quad (42)$$

where

$$\lambda_j=(r_{aj}+r_{bj})/R, \quad \mu_j=(r_{aj}-r_{bj})/R, \quad (43)$$

and

$$\rho=\{(\lambda_1^2-1)(\lambda_2^2-1)(1-\mu_1^2)(1-\mu_2^2)\}^{\frac{1}{2}}. \quad (44)$$

ϕ_1 and ϕ_2 denote the azimuthal angles of the two electrons, and δ is fixed as 0.75 throughout this calculation. The values of a_i^0 and E_0 are tabulated in Table I for $R=1.3$, 1.4, and 1.5 (atomic units). For the additional function, we have used two functions:

$$\varphi_1=\psi_1=\chi_1=(1/2\pi)\exp\{-\delta(\lambda_1+\lambda_2)\}$$
$$\times[\mu_1(\lambda_1^2-1)^{\frac{1}{2}}(1-\mu_1^2)^{\frac{1}{2}}\cos\phi_1$$
$$+\mu_2(\lambda_2^2-1)^{\frac{1}{2}}(1-\mu_2^2)^{\frac{1}{2}}\cos\phi_2],$$

$$\varphi_2=\psi_2=\chi_2=(1/2\pi)\exp\{-\delta(\lambda_1+\lambda_2)\}$$
$$\times[\mu_2(\lambda_1^2-1)^{\frac{1}{2}}(1-\mu_1^2)^{\frac{1}{2}}\cos\phi_1$$
$$+\mu_1(\lambda_2^2-1)^{\frac{1}{2}}(1-\mu_2^2)^{\frac{1}{2}}\cos\phi_2]. \quad (44)$$

The details of the calculation are quite analogous to those of the polarizability calculation[4] and we shall omit them here. The numerical values of the integrals required for our problem have been taken in part from the tables published by Kotani *et al.*[7] The tables made for the calculation of the polarizability have also been utilized. The results obtained are shown in Table II.

6. RESULTS AND DISCUSSION

In the preceding section, we obtained expressions for χ, σ, μ_r, and H_r, for fixed R and θ. We will now average these quantities over R for the lowest vibrational state. The method is the same as that used for calculat-

[7] Kotani, Amemiya, and Simose, Proc. Phys. Math. Soc. Japan 20, extra number (1938); M. Kotani and A. Amemiya, Proc. Phys. Math. Soc. Japan 22, extra number (1940).

TABLE III. The values a, b, c. Each figure in this table corresponds to that of Table II in the same arrangement. The units used are also the same as in Table II.

χ_{mol}	a	-3.5827	-0.7353	$+0.1952$	
	b	-1.2685	-1.4333	$+0.5086$	
	c	-1.26	-0.67	$+0.300$	
	$\langle\chi_{mol}\rangle_R=-3.6750-0.8168\sin^2\theta+0.2260\sin^2\theta$				(H_2)
	$\langle\chi_{mol}\rangle_R=-3.6469-0.7923\sin^2\theta+0.2167\sin^2\theta$				(D_2)
σ	a	$+2.773$	-0.0564	$+0.0683$	
	b	-1.199	$+0.9027$	-0.0317	
	c	$+0.300$	-0.385	-0.085	
	$\langle\sigma\rangle_R=2.731-0.0294\sin^2\theta+0.0667\sin^2\theta$				(H_2)
	$\langle\sigma\rangle_R=2.743-0.0372\sin^2\theta+0.0672\sin^2\theta$				(D_2)
$\Delta\mu_r$	a		-0.1677		
	b		-0.1978		
	c		$+0.110$		
	$\langle\Delta\mu_r\rangle_R=-0.1728$				(H_2)
	$\langle\Delta\mu_r\rangle_R=-0.1714M_p/M_d=-0.0857$				(D_2)
H_r	a	$+24.7689$		$+1.780$	
	b	-53.956		-3.372	
	c	$+77.000$		$+1.750$	
	$\langle H_r\rangle_R=24.783+1.689$				(H_2)
	$\langle H_r\rangle_R=(24.738+1.715)M_p/M_d$				(D_2)

ing the polarizability: we may express the R dependence of any quantity as

$$f(R)=a+b(R-R_e)+c(R-R_e)^2, \tag{45}$$

and average it by using the eigenfunction of the Morse-potential. (See Table III.)

The observed values are considered to be the averages of the above-obtained values over θ for the rotational states. As is well known, the rotational eigenfunctions can be expressed by spherical harmonics. It follows from the properties of the spherical harmonics that

$$\langle\cos^2\theta\rangle_{J,m_J}=\frac{2J^2+2J-1-2m_J^2}{(2J-1)(2J+3)},$$
$$\langle\sin^2\theta\rangle_{J,m_J}=\frac{2J^2+2J-2+2m_J^2}{(2J-1)(2J+3)}. \tag{46}$$

Taking into account the relation

$$\sum_{m_J=-J}^{+J} m_J^2=(1/3)(2J+1)(J+1)J, \tag{47}$$

we get

$$\langle\cos^2\theta\rangle_J=\tfrac{1}{3},\quad \langle\sin^2\theta\rangle_J=\tfrac{2}{3}. \tag{48}$$

The final results thus obtained are shown in Table IV, and indicate that our method is useful and gives good results even in the rough approximation. It should be noted that we have been able to avoid the difficulty in the perturbation treatment that the high-frequency terms become too large. Compared with the observed values, the calculated values of χ_{HF} and $\Delta\mu_r$ are large, and the high-frequency term of H_r is small. However, the errors arose to the same extent in the calculation of the polarizability when only three terms were added to the unperturbed wave function. Therefore, we may expect the situation to be improved by taking more terms. In our results, H_r of D_2 is smaller than $H_r/2$ of H_2. Experimentally the reverse is the case. This disagreement seems to come from the approximation in (45), which is not good for the first term of H_r.

To get more accurate results by our method, it is necessary (i) to adopt the more accurate wave function for Ψ_0, (ii) to increase the number of the added functions, and (iii) to carry out the calculations for a larger number of different values of R. Considerable improvement, however, may be expected by increasing the added functions to five or more.

It follows from (38), (39), (40), and (41) that

$$\chi=-2\sum a_i^0 a_j^0 \mathcal{K}_{ij}^{(4)} - (\alpha^2 a_0 R^2/8\mu_N)\sin^2\theta(\mu_r-\mu_N), \tag{49}$$

$$\sigma=(1/2)\sum a_i^0 a_j^0 \mathcal{K}_{ij}^{(5)} - (\alpha^2 a_0 R^2/4\mu_N)\sin^2\theta(\mu_N/R-H_r/2), \tag{50}$$

where

$$a_0=\hbar^2/me^2,\quad \alpha=e^2/\hbar c,\quad \text{and}\quad \mu_N=e\hbar/2Mc. \tag{51}$$

TABLE IV. Results of the calculations and comparison with experiment.

	H_2 Calculated	H_2 Observed	D_2 Calculated	D_2 Observed	Unit
χ_{mol}	-4.0689	-3.94, -3.99[a]	-4.0306		10^{-6} erg G^{-2} mole^{-1}
σ	2.756		2.763		10^{-6}
$\Delta\mu_r$	-0.1727	-0.11709 ± 0.00007[b]	-0.0857	-0.0594 ± 0.0030[e]	$e\hbar/2Mc$
H_r	26.472	27.32[d], 26.74 ± 0.05[e]	13.227	14.0 ± 0.6[d]; 13.43 ± 0.06[f]	Gauss
χ_{HF}	0.1507	0.0990 ± 0.001[b]; 0.093 ± 0.007[e]	0.1445		10^{-6} erg G^{-2} mole^{-1}
$\xi_\pm-\xi_0$	-3.922	-3.66 ± 0.20[b]; -3.9 ± 2.5[e]	-3.820	-5.1 ± 2.5[e]	10^{-31}/mole

[a] J. H. Van Vleck, *The Theory of Electric and Magnetic Susceptibilities* (Oxford University Press, Oxford, 1932), p. 279.
[b] N. J. Harrick and N. F. Ramsey, Phys. Rev. **88**, 228 (1951).
[c] N. F. Ramsey, Phys. Rev. **58**, 226 (1940).
[d] Kolsky, Rabi, and Ramsey, Phys. Rev. **57**, 677 (1940).
[e] Kolsky, Phipps, Ramsey, and Silsbee, Phys. Rev. **79**, 883 (1950).
[f] Kolsky, Phipps, Ramsey, and Silsbee, Phys. Rev. **80**, 483 (1950).

TABLE V. Comparison of our values χ_{mol} with those of other investigators.

	Hylleraas-Skavlem	Ramsey	Present calculation	Semi-empirical
Lamb's formula	3.1648	3.24		3.24
First-order correction	−0.5034	−0.63		−0.50
Sum	2.6614	2.61	2.711	2.74
Second-order correction	0.2864	0.07 (0.05*)	0.045	0.05
Total	2.9478	2.68 (2.66)	2.756	2.79

* The value obtained by the use of $H_r=26.74$ gauss.

On the other hand, Ramsey gives the formulas:[8]

$$\chi_{\text{Ramsey}}=\chi_0+\chi_{HF},\quad \chi_0=-\frac{\alpha^2 a_0}{6}\langle\Sigma|\mathbf{r}_j|^2\rangle,$$

$$\chi_{HF}=\frac{\alpha^2 a_0 R^2}{12}\left(1-\frac{\mu_r}{\mu_N}\right), \tag{52}$$

and

$$\sigma_{\text{Ramsey}}=\tfrac{2}{3}\alpha^2 a_0\left\langle\frac{1}{r_{a1}}-\frac{1}{4R}\right\rangle-\frac{\alpha^2 a_0 R^2}{6\mu_N}\left(\frac{\mu_N}{R^3}-\frac{H_r}{2}\right). \tag{53}$$

Putting $\sin^2\theta=\frac{2}{3}$, Eq. (49) agrees with (52) and the first term of (50) corresponds to the Hylleraas-Skavlem expression [Eq. (17) of reference 3]:

$$\tfrac{2}{3}\alpha^2 a_0\langle 1/r_{a1}-RZ_{a1}/2r_{a1}^3\rangle_{M}. \tag{54}$$

This, however, seems to be different from the corresponding term of (53). The explanation lies in Ramsey's assumption that "*the nuclei are at positions where the electric field is approximately zero.*" Under this assumption we obtain $\langle 1/4R\rangle=\langle Rz_{a1}/2r_{a1}^3\rangle$. The results of these calculations are shown in Table V. The difference between −0.63 (Ramsey) and −0.50 (Hylleraas-Skavlem) seems to be caused by Ramsey's assumption, and we list in column 4 the most appropriate values for the semiempirical σ value. The agreement with our values is quite excellent.

As to HD, the calculations become somewhat complicated because of the lower symmetry of this molecule. We will therefore publish it in a separate paper.

From the above calculation we see that if the Born-Oppenheimer (B-O) approximation holds, then $\mu_r=1$ and $\Delta\mu_r=0$; namely, electrons are completely separate from the nuclear system. This rather curious situation is most easily visualized if we consider the one-dimensional model. When a proton and an electron move in the x direction, then the correct wave function of the system may be described as

$$\exp\{ik(mx+MX)/(m+M)\}\cdot\chi(x-X),$$

where m,M, and x,X are the masses and coordinates of the electron and the proton, respectively. Then by the usual procedure we obtain the current due to both electron and proton.

On the other hand, if we use the B-O wave function instead of the correct one, namely

$$\exp\{ikX\}\cdot\chi(x-X),$$

we cannot obtain the electronic current moving with the proton. This consideration shows that the B-O approximation is not justified for the calculation of the electronic current moving within the nuclear framework.

For the H_2 molecule, however, the B-O approximation is adequate, as we know from experiment that electrons are almost completely separate from the nuclear system. The correction for the B-O approximation due to the incompleteness of the separability of variables introduces $\Delta\mu_-$, which is calculated by our formalism. We will consider the treatment of the other extreme case in a later paper.

In conclusion, the authors wish to express their sincere thanks to Professor M. Kotani for his guidance and to the members of his laboratory for their aid. They are also much indebted to Professor G. Araki for his discussions, and to Professor N. F. Ramsey for his kind interest. They are grateful to Mr. T. Kawai, Mr. S. Kawasaki, and Mr. S. Kuroda for their assistance in carrying out the numerical work. Part of this work was made possible by research grants from the Educational Ministry.

APPENDIX

Differentiating Eq. (19) with respect to H and taking the limit of vanishing H, μ and m_J, one obtains

$$\Sigma_j(\mathcal{H}_{ij}^{(0)}-S_{ij}E_0)(\partial a_j/\partial H)_0 - (\partial E/\partial H)_0\Sigma_j S_{ij}a_j^0=0. \tag{A-1}$$

Multiplying by $\Sigma_i a_i^0$ and using the relations

$$\Sigma_i(\mathcal{H}_{ij}^{(0)}-S_{ij}E_0)a_i^0=0, \tag{A-2}$$

$$\Sigma_{ij}S_{ij}a_i^0a_j^0=1,$$

we get

$$(\partial E/\partial H)_0=0$$

and therefore,

$$\Sigma_j(\mathcal{H}_{ij}^{(0)}-S_{ij}E_0)(\partial a_j/\partial H)_0=0. \tag{A-3}$$

Comparing (A-3) with (A-2), one can see that

$$(\partial a_j/\partial H)_0=pa_j^0. \tag{A-4}$$

Similarly, the derivatives of (19) with respect to μ and m_J give

$$(\partial a_j/\partial\mu)_0=qa_j^0 \tag{A-5}$$

and

$$(\partial a_j/\partial m_J)_0=ra_j^0, \tag{A-6}$$

[8] Ramsey's original form of this expression [Phys. Rev. 78, 703 (1950), Eq. (24)] is as follows:

$\sigma=(e^2/3\ mc^2)(0|\Sigma_k 1/r_k|0)-(\alpha^2 a_0 a^3/6\mu_N)(2Z\mu_N/a^3-\mu' H_r/MJ)$.

This is equivalent to Eq. (53) where some of the notation is altered for uniformity. The second and the third terms of Eq. (53) correspond to the second term of Ramsey's expression. This recombination is due to the different choice of gauge.

respectively. Thus we get:

$$a_j = a_j^0 + (\partial a_j/\partial H)H + (\partial a_j/\partial \mu)_0 \mu + (\partial a_j/\partial m_J)_0 m_J + \cdots = a_j^0(1 + pH + q\mu + r m_J + \cdots). \quad \text{(A-7)}$$

The normalization condition (17) gives

$$1 = S = \Sigma a_i^{0*} a_j^0 |1 + pH + q\mu + r m_J|^2 S_{ij} + H^2(\cdots) + H\mu(\cdots) + \mu m_J(\cdots) + \cdots. \quad \text{(A-8)}$$

From this we obtain

$$0 = (\partial S/\partial H)_0 = \Sigma a_i^{0*} a_j^0 S_{ij}(p^* + p);$$

i.e.,

$$p^* + p = 0,$$

which shows that p is purely imaginary. Similarly q and r are also shown to be purely imaginary. We can therefore write

$$a_j = a_j^0\{1 + i(\alpha H + \beta\mu + \gamma m_J) + \cdots\} = a_j^0 \exp\{i(\alpha H + \beta\mu + \gamma m_J)\} + (\text{second order terms}). \quad \text{(A-9)}$$

This shows that, if we multiply the wave function by an appropriate phase factor, the a's can be reduced to the form

$$a_j = a_j^0 + H^2 a_j^{(1)} + H\mu a_j^{(2)} + H m_J a_J^{(3)} + \mu m_J a_j^{(4)} + \cdots. \quad \text{(A-10)}$$

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Electron-Electron and Positron-Electron Scattering Measurements

ARTHUR ASHKIN,* LORNE A. PAGE,† AND W. M. WOODWARD
Laboratory of Nuclear Studies, Cornell University, Ithaca, New York

(Received November 27, 1953)

A counter experiment is described which measures the absolute differential electron-electron scattering cross section in the energy interval 0.6 to 1.2 Mev and the absolute differential positron-electron scattering cross section in the energy interval 0.6 to 1.0 Mev. The ratio of these cross sections is also measured with somewhat increased accuracy. The technique of measurement combines good resolution with large energy transfers between the particles to permit a sensitive test of the relativistic features of the Møller and Bhabha formulas. The results verify the Møller formula within the 7 percent experimental error. The Bhabha formula is verified within the 10 percent experimental error. The ratio of the Møller to the Bhabha formula is verified within about 8 percent experimental error.

INTRODUCTION

THE purpose of this report is to collect and summarize the results obtained at this laboratory on electron-electron and positron-electron scattering experiments.[1] The object of these experiments was to check the Møller and Bhabha formulas which are the predictions for e-e and p-e scattering, respectively, based on the Dirac theory. When these experiments were begun the only previous work in this field had been done with cloud chamber techniques.[2] Such experiments were not adequate to check properly either formula, since the relativistic features of the scattering begin to be appreciable only at large energy transfers between the incident and scattered particles, and these are rarely seen in the cloud chamber.

It is customary to discuss scattering cross sections in terms of angular distribution, since this is usually the experimentally measured quantity. However, our apparatus measures directly the energy transferred in the collision and this is probably a more meaningful concept in this particular problem. In any event the conservation laws provide a ready means of going from one to the other. We prefer to talk about the fraction v of the incident kinetic energy transferred in the collision.

The present experiments were designed to study the scattering at large energy transfers using a counter technique. A method has been devised which gives simultaneously a high resolution and good solid angle.

Apparatus was first designed for e-e scattering work, and extensive measurements were made in the energy range 0.6−1.7 Mev with $v=0.5$. Due to the inability to distinguish incident from scattered particles in e-e scattering, the Møller formula is symmetrical about $v=0.5$, the largest distinguishable energy transfer. Thus $v=0.5$ represents the most favorable situation for checking the Møller formula. In the energy range 0.6 −1.7 Mev the electrons are sufficiently relativistic to check the essential features of the Møller formula, which include, besides the Coulomb scattering, additional contributions arising from spin interactions. Since the first reports of this work showing agreement

* Now at Bell Telephone Laboratories, Murray Hill, New Jersey.
† Now at the University of Pittsburgh, Pittsburgh, Pennsylvania.

[1] Lorne A. Page, Phys. Rev. **81**, 1062 (1951); A. Ashkin and W. M. Woodward, Phys. Rev. **87**, 236 (1952).

[2] Ho Zah-Wei, Compt. rend. **226**, 1083 (1948); Groetzinger, Leder, Ribe, and Berger, Phys. Rev. **79**, 454 (1950); The most recent cloud chamber experiment reported is: G. R. Hoke, Phys. Rev. **87**, 285 (1952).

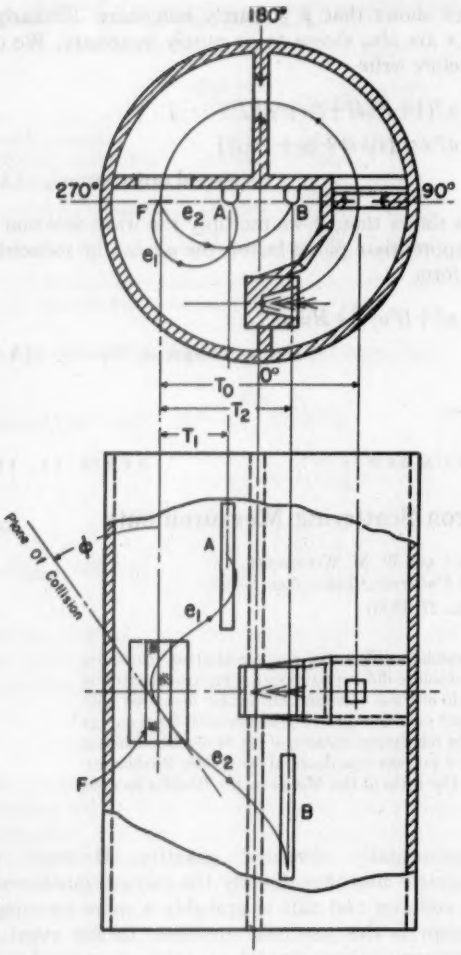

FIG. 1. Schematic view of the 270° scattering chamber used in electron-electron scattering measurements.

with the Møller formula,[1] several other experimenters, using electron accelerators,[3] have also found agreement at higher energies.

Following the work on *e-e* scattering, a new apparatus was built to handle the problem of *p-e* scattering. The positron-electron scattering shows many of the features of *e-e* scattering. The cross section is essentially Rutherford for small v, with deviations expected at large v. The cross section, however, is no longer symmetrical about $v=0.5$, since we have distinguishable particles. Also, in addition to the spin interactions found in the Møller formula, there exists the process called "virtual annihilation," or "exchange interaction." This is a contribution to the scattering cross section arising from the virtual annihilation and recreation of the positron and electron which go off in new directions. The net result is a scattering.

Since the first reports of this work showing agreement with the Bhabha formula,[1] a counter experiment was reported by Howe and MacKenzie[4] measuring the ratio of positron-electron to electron-electron scattering, giving good agreement with theory at 1.3 Mev.

APPARATUS

Geometry

Figure 1 gives a schematic view of the chamber used for the *e-e* scattering work. The chamber was evacuated to a mild vacuum and was surrounded by a solenoid and Helmholtz coils which provided a uniform magnetic field inside the tank. The top view shows the source placed at zero degrees. By a series of slits placed at 90° and 180°, an interval of momentum 12 percent wide was selected from the continuous β spectrum and brought to focus on the scattering foil F placed at 270°. Shown is an *e-e* collision taking place between one of the incident electrons and an atomic electron of the foil. The scattered particles are shown leaving the foil at some angles. They are bent in the magnetic field and are detected as a coincidence by two appropriately placed thin window Geiger counters A and B.

Since the horizontal displacement of the trajectory of the particle is determined only by the component of momentum perpendicular to the scattering foil and this is uniquely related to its kinetic energy, a counter placed parallel to the magnetic field will be sensitive only to particles of a particular energy. As one varies ϕ, the angle between the plane of collision and the vertical, the scattered particles travel along helical paths to different heights along counters A and B. Thus the fraction of all the scatterings observed with a particular v depends only on the length of counter used, and in principle can be made equal to unity with sufficiently long counters.

If, for example, one wants to study the differential scattering cross section for $v=0.4$, one places counter A at a distance from F such that it detects only electrons of energy $=0.4T_0$ where $T_0=$kinetic energy of the incident electron. By conservation of energy the other electron must have energy $0.6T_0$; therefore, B must be placed at a distance from F such that it detects only electrons of energy $0.6T_0$. Thus, for a fixed position of A there is a unique position for B. This property of the scattering we call the "coherence property." One can check its validity by studying the coincidence rate as a function of the position of B, for example, keeping A fixed. One expects to find a maximum at the position of coherence with a fall off on either side, the shape of which depends upon the particular geometry used. A typical curve is shown in Fig. 2.

Figure 3 gives a schematic view of the chamber used for the *p-e* scattering work. Here 180° rather than 270°

[3] Scott, Hanson, and Lyman, Phys. Rev. **84**, 638 (1951); Barber, Becker, and Chu, Phys. Rev. **89**, 950 (1953).

[4] H. A. Howe and K. R. MacKenzie, Phys. Rev. **90**, 678 (1953).

focusing is used in order to permit more effective shielding of the gamma rays from the positron source.[5] Actually, the 270° geometry is better in one respect, namely that the 90° and 180° slits define the beam which then passes cleanly through the foil at the 270° window, thus reducing slit scattering in the neighborhood of the scattering foil and resulting in low singles rates and correspondingly low backgrounds.

Another modification dictated by the need for adequate gamma-ray shielding was the so-called helical geometry. This refers to the fact that in the positron apparatus the source S is displaced downward from the horizontal plane containing the foil (see Fig. 3). In this situation the incident positron follows a helical path from the source to the foil, where it strikes with an upward component of momentum. As a result the collision products have an added upward component of motion which permits the shifting of the counters upward, thereby providing more room for shielding.

A consequence of this helical geometry is that all scattered particles of a given energy are not brought to focus on a vertical line, but rather on a curve situated roughly symmetrically about a vertical line. This deviation from the simple straight line results in a smearing out of the energy resolution somewhat beyond what would be expected solely from the widths of the counters and source.

The positron chamber can also be used to study *e-e* scattering by merely reversing the magnetic field and placing counter C_2 on the other side of the foil.

Counters

The Geiger counters used in the 270° chamber were made of ⅛-in. square tubing with 0.001-in. aluminum windows. The filling was 80 cm Hg of helium with 1 cm Hg of ethyl alcohol. The Geiger counters used in the 180° chamber were made of 0.8-in. square tubing with windows prepared by evaporating aluminum on

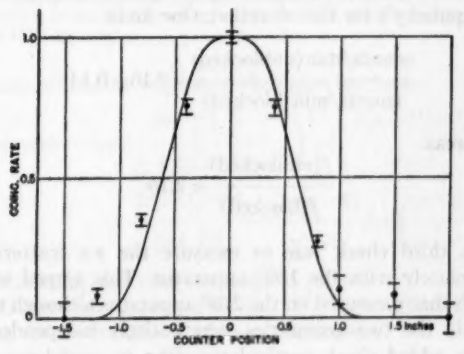

FIG. 2. A typical coherence curve.

[5] The Co^{56} positron sources which were used for reasons of energy and lifetime have approximately three energetic gamma rays for each positron. The $Sr^{90}-Y^{90}$ electron source is free of gamma rays.

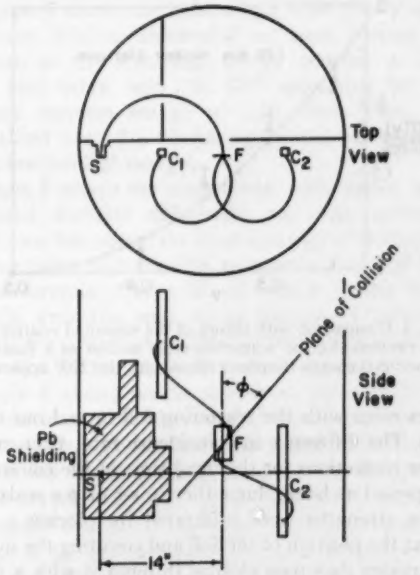

FIG. 3. Schematic view of the 180° scattering chamber used for both positron-electron and electron-electron scattering measurements.

0.0005-in. Mylar.[6] The filling was 7.5 cm Hg of argon and 1 cm Hg of ethyl alcohol. These counters, which had superior counting characteristics to the high-pressure ones, suffered from the disadvantage that they had to be filled each time after the chamber was evacuated.

Scattering Foils

With the 270° chamber two foils were used, 0.5 mm/cm² of collodion and 4.5 mg/cm² of beryllium. With the 180° chamber Mylar foils[6] of thickness 0.9 mg/cm² and 1.7 mg/cm² were used.

Source Strengths

The *e-e* scattering work using the 270° chamber was done with about 10 mc of $Sr^{90}-Y^{90}$. The *e-e* scattering work using the 180° chamber was done with about 2.5 mc of $Sr^{90}-Y^{90}$. The *p-e* scattering work was done with about 1 mc of Co^{56}.[7]

EXPERIMENTAL PROCEDURE

With the counters located for the desired energy sharing and the magnetic field set for the desired incident energy, one measures the coincidence rate and

[6] Mylar is a plastic film supplied by the DuPont Company. It is a polymer containing hydrogen, carbon, and oxygen.

[7] We are indebted to Professor Martin Deutsch, Massachusetts Institute of Technology; Professor P. V. C. Hough, University of Michigan; and Professor A. J. Allen, University of Pittsburgh for kindly providing us with the cobalt sources used in this work We also would like to thank Professor Norman Bonner of Cornell University for assistance with the chemistry of the source preparation.

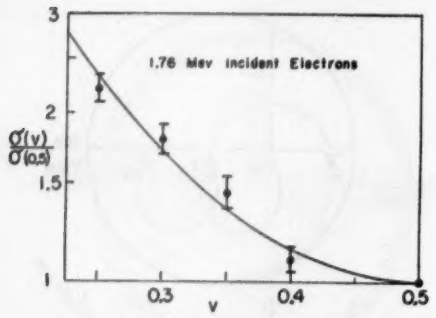

FIG. 4. Comparison with theory of the measured relative differential electron-electron scattering cross section as a function of the fractional energy transfer v taken with the 270° apparatus.

singles rates with the scattering foil in and out of the beam. The difference in coincidence rate, after making proper corrections for the dead time of the counters, is interpreted as being due either to e-e or p-e scattering. Source strengths were calibrated by placing a small hole at the position of the foil and counting the number of particles that pass cleanly through it with a Geiger counter placed some distance behind the hole. This measurement gives the source strength at one point on the foil. It is necessary to scan horizontally and vertically across the foil to determine the variation of the incident particle intensity with position on the foil.

Knowing the composition and weight of the foil, the above data on coincidence rate and absolute source strength is sufficient to determine the absolute differential cross section for e-e and p-e scattering. Actually, one must compute a geometrical factor before one can interpret the coincidence rate in terms of a cross section. This factor we call f, the efficiency against vertical loss. It is computed on the basis of the detailed orbits of the scattered particles from different parts of the scattering foil. This factor is a measure of the fraction of all the scatterings of a particular energy sharing which are detected by the counters. It differs from unity because of the restricted length of the counters.[8] As we saw before, this restricts the range of ϕ's which can be detected. This factor takes into account the variation in efficiency with which different parts of the foil are counted. It includes any variations in the number of particles incident at different points on the foil and the variation of the cross section with position on the foil. In the 180° apparatus it includes the smearing out of the energy resolution due to the helical geometry.

The accuracy of the vertical loss factor was checked experimentally in several ways.

In taking data at a particular energy sharing between the particles, for example (0.60−0.40), one can use two possible arrangements of counter positions. One can place the upper counter at a position where it detects particles of energy $0.4T_0$ and the lower at $0.6T_0$. Alternatively, one could place the upper at 0.6 and the lower at 0.4. Either way we should measure the same cross section. This implies that the ratio of the observed counting rates should be equal to the ratio of the computed f's for the two situations. One finds

$$\frac{\text{counts/min}(0.60-0.40)}{\text{counts/min}(0.40-0.60)}=1.55\pm0.09.$$

The computed ratio of the f's is

$$\frac{f(0.60-0.40)}{f(0.40-0.60)}=1.53.$$

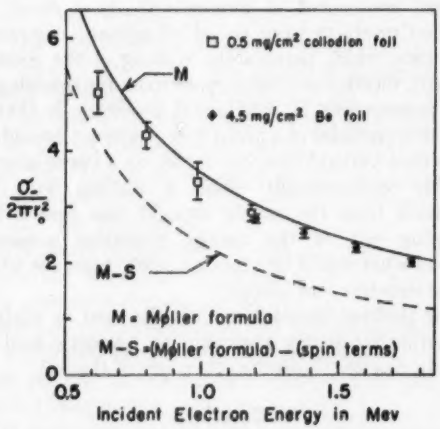

FIG. 5. Comparison with theory of the measured absolute differential electron-electron scattering cross section at $v=0.5$ as a function of the kinetic energy of the incident electron taken with the 270° apparatus.

Another check is to change the actual length of the upper counter in the situation (0.60−0.40). The ratio of the counting rates should again equal the ratio of the computed f's for this situation. One finds

$$\frac{\text{counts/min(unblocked)}}{\text{counts/min(blocked)}}=2.16\pm0.14,$$

whereas

$$\frac{f(\text{unblocked})}{f(\text{blocked})}=2.16.$$

A third check was to measure the e-e scattering absolutely with the 180° apparatus. This agreed well with that measured on the 270° apparatus although the f's in the two geometries were entirely independent. This added check particularly gives us confidence in our understanding of the apparatus.

Care was taken to avoid errors due to multiple scattering in the scattering foil. A rather sensitive indication of the presence of multiple scattering is the

[8] The counter lengths had to be restricted for various reasons, such as having to have both counters in the same position, shielding, and the size of the useful magnetic field.

broadening of the coherence curve beyond the computed width. Another check is to study the counting rate as a function of the thickness of the scattering foil wherever possible.

RESULTS

The Møller formula[9] can be written in the following form:

$$\sigma_M(\gamma,v)dv=2\pi r_0^2\frac{\gamma^2}{(\gamma-1)^2}\frac{dv}{(\gamma+1)}\times\left[x^2-3x+\left(\frac{\gamma-1}{\gamma}\right)^2(1+x)\right],$$

where

$$r_0=e^2/mc^2,\quad \gamma=E_{\text{total}}/mc^2,$$
$$E_{\text{total}}=mc^2+T,\quad x=[v(1-v)]^{-1}$$

and v is the fractional energy transfer. The Bhabha

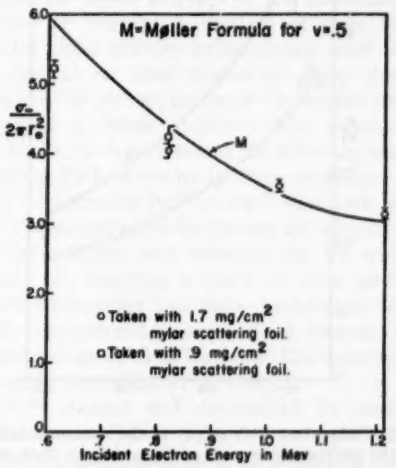
M=Møller Formula for v=.5
M
Taken with 1.7 mg/cm² mylar scattering foil.
Taken with .9 mg/cm² mylar scattering foil.
Incident Electron Energy in Mev

FIG. 6. Comparison with theory of the measured absolute differential electron-electron scattering cross section at $v=0.5$ as a function of the kinetic energy of the incident electron taken with the 180° apparatus.

formula[10] can be written in the following form:

$$\sigma_B(\gamma,v)dv=2\pi r_0^2\frac{dv}{(\gamma+1)}\left[\frac{1}{v^2}\left(\frac{\gamma}{\gamma-1}\right)^2-\frac{1}{v}\frac{2(\gamma+1)^2-1}{\gamma^2-1}+\frac{3(\gamma+1)^2+1}{(\gamma+1)^2}-v\frac{2\gamma(\gamma-1)}{\gamma^2+1}+v^2\left(\frac{\gamma-1}{\gamma+1}\right)^2\right],$$

where r_0, γ, and v are the same as above.

If one omits the contributions of the virtual annihilation terms, the Bhabha formula becomes

$$\sigma_{B-A}(\gamma,v)dv=2\pi r_0^2\frac{dv}{(\gamma+1)}\left[\frac{1}{v^2}\left(\frac{\gamma}{\gamma-1}\right)^2-\frac{1}{v}\left(\frac{\gamma+1}{\gamma-1}\right)+\frac{1}{2}\right].$$

[9] C. Møller, Ann. Physik 14, 531 (1932).
[10] H. J. Bhabha, Proc. Roy. Soc. (London) A154 195 (1936).

Figure 4 shows the comparison with theory of the measured relative differential e-e cross section as a function of the fractional energy transfer v. These data were taken with the 270° apparatus for fixed incident electron energy of 1.76 Mev. They were normalized to $v=0.5$. Similar agreement was found for 1.15-Mev incident energy.

Figure 5 shows the comparison with theory of the measured absolute differential e-e cross section at $v=0.5$ as a function of the kinetic energy of the incident electron taken with the 270° apparatus. Curve M is the Møller formula. Curve $M-S$, which is the Møller formula with the spin terms deleted, is drawn to indicate the magnitude of the spin terms. The preference of the data for the Moller formula is apparent.

Figure 6 shows results similar to those of Fig. 5, only this time taken with the 180° apparatus. We notice that the two high-energy points taken with the 1.7-mg/cm² foil are in good agreement with theory. The 0.82-Mev point with this foil thickness lies a little low, presumably because of multiple scattering, since a reduction in the foil thickness to 0.9 mg/cm² improves the agreement with theory. The point at 0.61 Mev is low, presumably because of multiple scattering, even with the 0.9-mg/cm² foil.

In Fig. 7 is displayed the measured energy dependence of the absolute positron-electron cross section for $v=0.5$ in units of $2\pi r_0^2$ taken with the 180° apparatus. Curve B is the prediction of the Bhabha formula. Curve $B-A$ is the prediction of the Bhabha formula less the annihilation term referred to above. As in Fig. 6 for the absolute electron measurements, the agreement with theory is good at the higher energies, but the value at 0.61 Mev is somewhat low, presumably because of multiple scattering. There seems to be little doubt that the

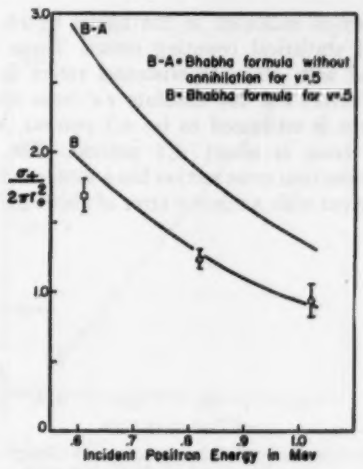

FIG. 7. Comparison with theory of the measured energy dependence of the absolute differential positron-electron scattering cross section for $v=0.5$, taken with the 180° apparatus.

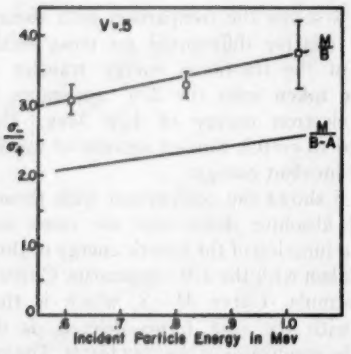

FIG. 8. The ratio of the differential electron-electron to positron-electron cross section, for $v=0.5$, taken with the 180° apparatus. This combines the results of Figs. 6 and 7.

measurements fit the Bhabha curve much better than the $B-A$ curve.

In Fig. 8 we have plotted the ratio of the $e-e$ to the p-e cross section as measured with the 180° apparatus. This combines the results of Figs. 6 and 7. The comparison of the data with the ratio of the Møller to Bhabha formulas and the Møller to Bhabha less the annihilation term is also shown. The experimental points are in quite good agreement with the M/B curve. Even the 0.61-Mev point lies on the theoretical curve, which serves to substantiate the supposition that both absolute numbers for the cross section at this energy were low because of multiple scattering.

Figure 9 shows the measured ratio of electron-electron to positron-electron cross section as a function of v for a fixed incident energy of 0.61 Mev. Curves M/B and $M/B-A$ mean the same as before. The experimental evidence again favors M/B over $M/B-A$.

ERRORS

The errors indicated in the above figures are the standard statistical counting errors. These must be combined with other experimental errors to give an over-all error. For the absolute e-e cross section the total error is estimated to be ±7 percent, while the relative error is about ±5 percent. The absolute positron-electron cross section has a total error of about ±10 percent with a relative error of about ±7 percent. This latter statement is not true, of course, at the points where the positron counting errors approach ±10 percent.

The measurement of the ratio of the Møller to the Bhabha formula has an error of about ±8 percent.

The errors to be associated with the e-e and p-e cross sections as a function of v are the relative errors of ±5 percent and ±7 percent, respectively, except at a few points where the statistics are a little worse.

CONCLUSIONS

The Møller formula is verified in the energy interval 0.6 to 1.2 Mev to an accuracy of about 7 percent standard deviation. The existence of the spin terms is established.

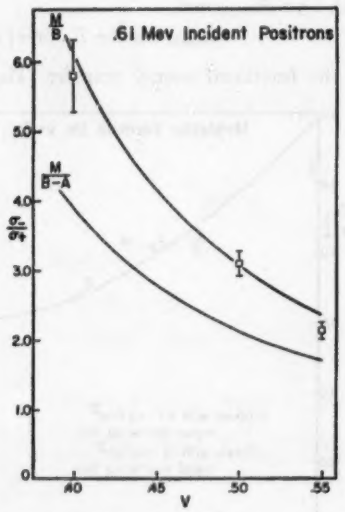

FIG. 9. Comparison with theory of the measured ratio of the differential electron-electron to positron-electron cross section as a function of v for an incident energy of 0.61 Mev.

The Bhabha formula is verified in the energy interval 0.6 to 1.0 Mev to an accuracy of about 10 percent standard deviation. The existence of the virtual annihilation term is established.

The ratio of Møller to Bhabha formulas is verified to within about 8 percent experimental error.

Half-Lives of Pu^{240} and Pu^{239}†

G. W. FARWELL,* J. E. ROBERTS,‡ AND A. C. WAHL§
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

(Received December 22, 1953)

The half-lives of Pu^{240} and Pu^{239} have been determined by measuring the specific alpha activity and the Pu^{240} content of four plutonium samples which had received different amounts of neutron irradiation. With the assumption that Pu^{238} contributes 13 percent to the observed increase in alpha activity, the half-life of Pu^{240} is 6300±600 years. The half-life of Pu^{239} is found to be 24 400±500 years.

THE specific alpha activity and Pu^{240} content of four plutonium samples have been measured. All the plutonium samples were prepared from uranium which had been neutron-irradiated in a graphite lattice pile. The chemical procedure used to separate the plutonium from the uranium and fission products and to purify it is capable of removing all other elements from plutonium.

The specific alpha activity of the plutonium was determined as follows: A solution of pure Pu was assayed for alpha activity by mounting small aliquots (~0.25 μg Pu) on glass microscope cover slips and counting in a "50 percent geometry" ionization chamber connected to a linear amplifier. The solution was assayed for Pu by at least one of the following methods: (1) weighing PuO_2 made by igniting an aliquot of the solution; (2) titrating Pu^{III} to Pu^{IV} with ceric sulfate. The Pu^{III} was prepared by fuming an aliquot of the solution with H_2SO_4 and reducing the Pu with zinc amalgam[1]; (3) weighing a piece of pure plutonium metal before dissolving it to make the solution. Samples *A* and *B* were assayed by method (1). Samples *C* and *D* were assayed by all three methods. The three methods of assay gave very satisfactory checks.

The Pu^{240} content was determined by measuring the spontaneous fission rates of several ~0.5-mg samples of Pu, each mounted in an ionization chamber attached to a battery-operated linear amplifier. The Pu was mounted in a thin film on a platinum disk. The effective weight of Pu^{239} was determined by placing the chamber containing the Pu in a known slow neutron flux and measuring the neutron-induced fission rate. From previous calibrations against the mass spectrograph results the spontaneous fission rate was converted to weight of Pu^{240}.

TABLE I. Specific alpha activity of plutonium samples vs Pu^{240} content.

Sample No.	Specific alpha activity (counts/min per μg Pu)	Pu^{240} content (μg Pu^{240} per μg Pu)
A	69 480±100	0.00063±0.00006
B	69 960±100	0.0034 ±0.0002
C	72 510±100	0.0133 ±0.0006
D	72 930±100	0.0157 ±0.0006

The data are given in Table I and plotted in Fig. 1.

If all of the increase in the specific alpha activity is due to the presence of Pu^{240}, then the alpha activity of 0.01 μg of Pu^{240} is

$$[71\,650-(69\,240)(0.99)]\ \text{counts/min}$$
$$=3100\ \text{counts/min}.$$

Then if one assumes 51 percent geometry for the 50 percent geometry chamber, the specific alpha activity of Pu^{240} is 6.08×10^6 dis/min per μg Pu^{240} and the half-life of Pu^{240} is 5500 years.

However, according to Perlman and Seaborg,[2] 13 percent of the increase in specific alpha activity is due to the presence of Pu^{238}. This correction increases the result for the half-life of Pu^{240} to 6300 years. The estimated probable error for the half-life of Pu^{240} is 10 percent.

From the intercept of the specific alpha activity curve the half-life of Pu^{239} can be calculated. Assuming 51 percent geometry for the 50 percent geometry chamber, the specific alpha activity of Pu^{239} is 136 000

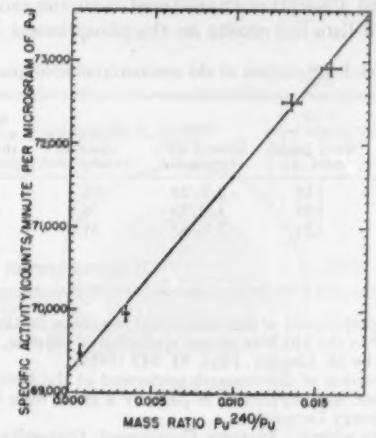

FIG. 1. Specific alpha activity of Pu^{240}–Pu^{239} mixtures vs mass ratio Pu^{240}/Pu for the mixtures.

† This paper is, with minor literary changes, a report dated October 11, 1945, of work done at the Los Alamos Scientific Laboratory of the University of California. It has recently been declassified. The work was performed under a contract for the Manhattan Engineer District.

* University of Washington, Seattle, Washington.

‡ University of California, Radiation Laboratory, Livermore, California.

§ Washington University, St. Louis, Missouri.

[1] We are indebted to Mr. Harold Boaz of group CM-9 for developing this method and performing the titration.

[2] I. Perlman and G. T. Seaborg, Los Alamos Report MUC-GTS-1872, July 28, 1945 (unpublished).

dis/min per μg of Pu^{239}, and the half-life of Pu^{239} is 24 400 years. The estimated probable error for the half-life of Pu^{239} is 2 percent, the uncertainty coming mainly from the uncertainty in the geometry.

We wish to express our thanks to Dr. Emilio Segrè for his help and encouragement during the course of the work and for his continued interest which made possible the publication of the results.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Evidence for Si^{32}, a Long-Lived Beta Emitter*†

ANTHONY TURKEVICH, *Institute for Nuclear Studies, University of Chicago, Chicago, Illinois, and Chemistry Division, Argonne National Laboratory, Lemont, Illinois*

AND

ARTHUR SAMUELS,‡ *Institute for Nuclear Studies, University of Chicago, Chicago, Illinois*

(Received August 24, 1953)

Neutron-irradiated quartz has been found to contain 14.3-day P^{32} more than two years after the end of irradiation. This is evidence for a long-lived Si^{32} formed from stable Si^{30} by the capture of two neutrons. The ratio of half-life of Si^{32}, in years, to neutron capture cross section of Si^{31}, in barns is 600.

PREVIOUS work[1,2] has indicated that the unknown nuclide Si^{32} might be a long-lived beta emitter. This communication reports direct evidence (via isolation of its P^{32} daughter) for such a long lived Si^{32}.

Quartz, that had been intensively irradiated with thermal neutrons at the Hanford pile and allowed to cool for more than two years,[3] has been found to contain small amounts of 14.3-day P^{32}. The radiophosphorous was separated from approximately 100-gram samples of the quartz by volatilization of the silicon as SiF_4 from sulfuric and hydrofluoric acids in the presence of phosphate carrier. It was then purified by ammonium phosphomolybdate and magnesium ammonium phosphate precipitations. Included also were CuS scavengings and decontamination from vanadium by reduction of added V^{5+} carrier prior to some of the phosphate precipitations. The identification of the separated radioactivity as P^{32} was made from the decay and absorption characteristics of the radiations.

Table I lists the results on the phosphorous activity isolated from three samples of the quartz. The SiF_4 evaporation took up to two weeks. The carrier was added in several portions during the evaporation. Column three of the table lists the midpoint of this evaporation period. Column four gives the observed initial counting rate of the purified phosphorous. The chemistry extended over a period of 10 to 26 days and the chemical yields were between 15 and 50 percent. Finally, the last column lists the P^{32} activity of the original quartz calculated from the observed activity, the chemical yields, and the decay periods. As an additional check on radiochemical purity, the decay of the phosphorous sample IV was followed for 10 days and then subjected to an additional cycle of radiochemical purification. The result was no change (less than 5 percent) in specific activity and in the absorption characteristics.

TABLE I. P^{32} content of old neutron-irradiated quartz.

Sample	Wt of quartz used (g)	Date of SiF_4 evaporation	P^{32} activity observed (counts/min)	Specific activity of the quartz (counts/min g)
II	110	2/6/53	6.4	0.5
III	120	3/2/53	6.9	0.5
IV	120	5/6/53	31.0	0.8
			Weighted average	0.66

Several samples of unirradiated quartz, treated chemically in an identical manner, showed no P^{32} activity. Our limit of detection was about one-fiftieth of that observed in the most active sample isolated from the irradiated quartz.

Table I shows that there is P^{32} activity associated with neutron-irradiated quartz long after that formed directly by pile radiations has died away. In addition this activity is essentially constant over a period of about three months. A reasonable interpretation is that a long-lived parent, Si^{32}, has been formed from stable Si^{30} by the capture of two neutrons. From the weighted average of the specific activities indicated in Table I (0.66), the detection efficiency of our end-window proportional counters for P^{32} radiations ($\sim$45 percent), the irradiation conditions (flux and time), the thermal neutron capture cross section of Si^{30} (0.2 barn) and the half-life of Si^{31} (156 min), we calculate the ratio of half-life of Si^{32}, in years, to neutron capture cross section of Si^{31}, in barns, to be 600. The thermal neutron absorption cross section of Si^{31} is not known. For a cross section of 0.1 barn, the half-life of Si^{32} is calculated to be 60 years.

* Since preparation of this manuscript, evidence for the production of Si^{32} in the 340-Mev proton spallation of chlorine, has been presented by M. Lindner, Phys. 91, 642 (1953).

† The portion of this research performed at the University of Chicago has been supported, in part, by a grant from the U. S. Atomic Energy Commission.

‡ Present address: Anatomy Department, University of Colorado Medical School, Denver, Colorado.

[1] M. Lindner, Phys. Rev. 89, 1150 (1953).

[2] A. Turkevich and Althea Tompkins, Phys. Rev. 90, 247 (1953).

[3] We thank Dr. F. T. Hagemann of the Argonne National Laboratory for making this quartz available to us.

Neon-18†

J. D. Gow and Luis W. Alvarez
Radiation Laboratory, Department of Physics, University of California, Berkeley, California

(Received January 8, 1954)

A new radioactivity which we assign to the nucleus Ne^{18} has been investigated. The decay is by the emission of positrons with an upper limit of 3.2±0.2 Mev and a half-life of 1.6 seconds. The spectrum of β rays was compared directly in a magnetic spectrograph with the known radiation of He^6. The β-ray energy and decay rate correspond to a log ft value of 2.9±0.2, placing the activity in the superallowed class.

INTRODUCTION

STUDIES of the mirror nuclei and of the isobaric triads have contributed greatly to the understanding of nuclear forces, and to the clarification of the theory of beta decay. Until the finding of Ne^{18}, which is described in this paper, only six sets of isobaric triads were known. These are $(Li-Be-B)^8$,[1] $(Be-B-C)^{10}$,[2] $(B-C-N)^{12}$,[3] $(C-N-O)^{14}$,[2] $(F-Ne-Na)^{20}$,[1] and $(Na-Mg-Al)^{24}$.[4] Be^6 is certainly proton unstable, and F^{16} has been shown to be proton unstable.[1] Both of these instabilities are predicted on the simple assumption of charge symmetry of nuclear forces plus the calculated Coulomb energy differences of the extreme members of the triads, plus the known masses of He^6 and N^{16}. According to the same considerations, Ne^{18} should be radioactive with a period of several seconds, and emit positrons with an upper limit of about 3.3 Mev.

Since one can presumably predict the properties of Ne^{18} with some confidence, it would appear at first sight that Ne^{18} should be very easy to find in the laboratory. Actually, it has turned out to be very difficult to observe Ne^{18} in the omnipresent background of Ne^{19} and N^{16}. These two activities are made whenever fluorine is bombarded with protons energetic enough to produce the reaction $F^{19}(p,2n)Ne^{18}$. They both have very short periods, but unfortunately the period of Ne^{18} is shorter than that of either of the impurities, so there is no known way to separate the wanted activity from the unwanted. Several attempts were made to observe the growth of F^{18} from an unobservable Ne^{18}, mixed with Ne^{19}. This is the method of the "rayless transition," which was used by Rutherford in the early days of radioactivity. Liquid C_7F_{16} was bombarded with protons, and a stream of He gas carried Ne^{19} through active charcoal at liquid-air temperature into a large glass bottle containing a thin-walled Geiger counter. Neon, helium, and hydrogen are the only gases that can pass through cold charcoal, and a pure 18.5-second period (Ne^{19}) of high intensity was observed in the bottle. It was hoped that by observing the activity of the F^{18} daughter as a function of the delay time in filling the bottle the lifetime of Ne^{18} could be measured. These experiments were unsuccessful. This proved to be due to a long "holdup" time associated with the passage of neon gas through the charcoal trap. Although the volume of the trap was quite small, delay times for Ne^{19} as long as 40 seconds were measured. Ne^{18} was finally found by the use of a beta-ray spectrograph. This was possible because its end point is slightly higher than that of Ne^{19}, and the positrons from the decay of N^{16} to the "pair state" in O^{16} contribute a very small background.

EXPERIMENTAL TECHNIQUES

After some preliminary experiments with a very crude beta-ray spectrograph, which had indicated the presence of a short-lived Ne^{18}, we were able to use a much finer instrument, through the courtesy of Dr. Roger Wallace. His 180° magnetic spectrograph is patterned after that of Lawson and Tyler[5] and uses two proportional counters in coincidence as the detecting device. With this instrument, it was possible to separate completely the radiations of the 18.5-second Ne^{19} from those of the new shorter period, on the basis of beta-ray energy.

The short-period activity was induced by proton bombardment of targets of teflon $(CF_2)_N$ and a clear

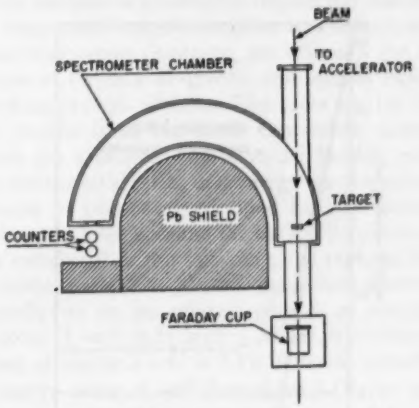

FIG. 1. Schematic experimental setup.

† This work was performed under the auspices of the U. S. Atomic Energy Commission.

[1] L. W. Alvarez, Phys. Rev. **80**, 519 (1950).
[2] Sherr, Mueher, and White, Phys. Rev. **75**, 282 (1949).
[3] L. W. Alvarez, Phys. Rev. **75**, 1815 (1949).
[4] A. C. Birge, Phys. Rev. **85**, 753 (1952).
[5] J. L. Lawson and A. W. Tyler, Rev. Sci. Instr. **11**, 6 (1940).

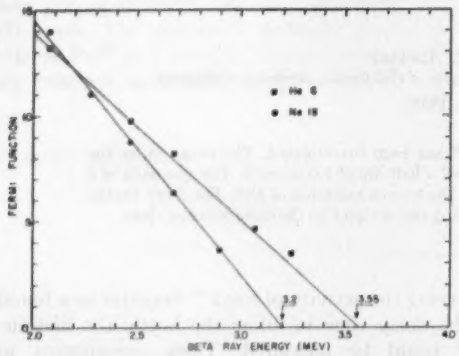

FIG. 2. Kurie plots for He^6 and Ne^{19}.

crystal of LiF. The target was bombarded by protons from the linear accelerator, and it was permanently fixed in the "source position" of the spectrograph (Fig. 1). Intensity measurements of the positron spectrum were made while the beam was turned off, after short bombardments of the target. The beam was monitored by a novel scheme due to Dr. W. K. H. Panofsky. The Faraday cup, which caught the beam after it passed through the target, was connected to a parallel *RC* network with a time constant equal to the mean life of the radioactivity. A fast recording electrometer gave a record of the voltage across the *RC* circuit. The reading of the electrometer at the instant the beam is turned off is obviously proportional to the activity of the target. The advantage of this system is that variations of the beam intensity during the bombardment do not affect the proportionality of integrator reading and source activity.

The beta spectrum could be investigated only above the end point of the strong Ne^{19} activity of 2 Mev. The shape of the spectrum between 2 Mev and 3.3 Mev was not appreciably affected by the LiF target thickness of 1.5 mm. LiF was very convenient to use since it

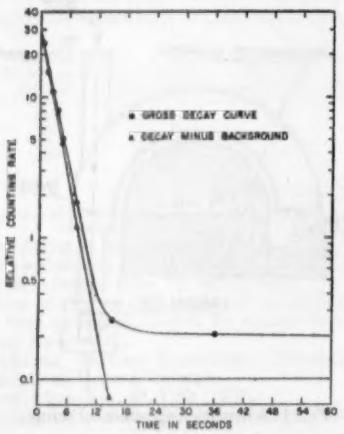

FIG. 3. Decay curve for Ne^{19}. Half-life = 1.6 sec.

yields the well-known 0.85-second period of He^6, by the $Li^7(p,2p)He^6$ reaction. Its half-life and its upper limit of 3.55 Mev are very close to the corresponding properties of Ne^{19}, so no absolute calibration of the spectrograph was necessary. To observe He^6, one had merely to reverse the magnet current and change the time constant of the integrator circuit.

The output pulses of the spectrometer counters were fed to pulse shapers and to a conventional coincidence circuit. The coincident pulses were fed to a scaler with adjustable scaling factor, and the scaled output pulses were recorded on a Brush tape recorder. In order to prevent variations in the time after bombardment at which counting began, the pulses to the Brush tape recorder passed through a relay. The relay closed when the flip gates which served to interrupt the beam were in place. This arrangement gave a highly reproducible counting cycle.

The procedure for taking the data was to set the spectrometer magnet current to a desired value, open the flip gate for the desired length of bombardment, close the gate, and allow the decay to be recorded for long enough to establish a good background value. Enough runs were made in this manner at each value of the magnet current to give adequate statistical accuracy. The reading of the recording electrometer at the end of the bombardment was associated with each tape record of a decay.

The data were analyzed by adding all the runs at a given magnet current and plotting the resultant data as a function of time. The background was subtracted to get the net counts of the desired activity in the first second after bombardment. This number was divided by the total of the individual electrometer readings to give a normalized "count in first second" for each value of magnet current.

The data so obtained for He^6 were arranged in the form of a Fermi plot against an energy scale calculated from the magnet current and adjusted to give the tabulated end point of 3.55 Mev. The data for the 1.6-sec positron activity were plotted on the same scale (Fig. 2). The best fit by eye for these data yields an end point of 3.2 Mev.

A decay curve taken at the magnet current that gave the best ratio of 1.6-sec activity to background is shown in Fig. 3. From this we assign a value of the half-life of 1.6 ± 0.2 sec.

NUCLEAR ASSIGNMENTS

The activity was shown to be characteristic of fluorine by its appearance in proton-bombarded targets of LiF and CF_2, and its absence in targets of CH_2. Ne^{20*}, which is the compound nucleus formed in the reaction $F^{19}+p$, cannot decay by nucleon emission to any nucleus having either N or Z higher than 10. Since the resultant nucleus which we have investigated is positron-active, it lies on the neutron-deficient side of the stability line. The only presently unknown nuclear

species which could be produced by 30-Mev protons on fluorine are Ne^{18}, Ne^{17}, and F^{16}. Since F^{16} is known to be unstable,[1] the activity must be either Ne^{18} or Ne^{17}. If one assumes the activity to be Ne^{17}, the beta-ray energy, coupled with the known mass of F^{17}, gives a minimum possible mass for Ne^{17}. From this value, and the known masses of the other particles involved, one can calculate the threshold energy for the reaction $F^{19}(p,3n)Ne^{17}$. The threshold would be 25.8 Mev. Although the backgrounds and low counting rates prevented the establishment, with high precision, of the threshold for producing the Ne^{18} activity, the activity was solidly in evidence when the incident proton energy was reduced to 24 Mev. Since this energy is below the minimum possible energetic threshold for producing N^{17}, we can eliminate that possibility. Even in the absence of this convincing evidence, one could feel sure that Ne^{17} would have much more energetic positrons (its triad counterpart is N^{17}), and would assign the new activity to Ne^{18} on the grounds that it behaves just as the theory would predict that it should.

CONCLUSION

Our value of 3.2 Mev for the upper limit of the beta-ray spectrum, coupled with the lifetime of 1.6 sec, gives a log *ft* value for the decay of Ne^{18} of 2.9 ± 0.2. This value clearly places the decay character in the same class as the other known $A=4n+2$ nuclei. The only other known nucleus with such a highly allowed beta decay is He^6, which has a log *ft* value of 2.95. The close agreement between the calculated mass of Ne^{18} and the observed upper limit of the positrons indicates strongly that the transition goes to the ground state of F^{18}.

ACKNOWLEDGMENTS

We would like to express our appreciation to Dr. Roger Wallace for much valuable assistance, and to Dr. Sumner Kitchen, who also assisted in the experimental program. Our thanks also go to the members of the linear-accelerator operating group who aided the experiment in many ways.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Photon Splitting in a Nuclear Electrostatic Field*

MARK BOLSTERLI†
Washington University, St. Louis, Missouri
(Received December 18, 1953)

The cross section for the splitting of a photon into two photons in a nuclear electrostatic field has been calculated from the vacuum polarization Hamiltonian of Euler and others to first order in $e^2/\hbar c$ for low-energy incident photons ($\rho<mc$). For a favorable experimental case, photons of energy 840 kev incident on lead with antiparallel product photons each emitted at 90° relative to the incident photon, the cross section is 2.3×10^{-33} cm²/sterad².

THE nonlinear terms in the Maxwell equations arising from the polarization of the vacuum[1] result in several interesting effects: coherent photon scattering by a nuclear electrostatic field,[2] scattering of photons by photons,[3] and photon splitting into two product photons in a nuclear electrostatic field, the last briefly discussed by Williams.[4] The cross sections for the first two processes are quite small and hard to verify experimentally, the first because of the difficulty in separating out the coherent scattering by the nuclear electrostatic field from other coherent nuclear and atomic-electron photon scatterings (however, see Wilson[5]), the second because of the lack of gamma-ray sources capable of furnishing enough photons. The cross section for the third process is no less small, but strong gamma-ray sources are available, high-Z nuclei furnish relatively large electrostatic fields, and energy discrimination can be made to eliminate unwanted inelastic scatterings which act as a background for the splitting process.

As calculated in the following, the cross section for the production of two photons oppositely directed and perpendicular to an original photon of energy 1.65 (in units of mc^2—0.84-Mev γ of Mn^{54}) incident on a nucleus of charge $Z=82$ is 2.3×10^{-33} cm²/sterad². (We use energy units of mc^2 throughout.) Of the product photons, nearly all have energy between 0.4 and 1.3

* Assisted by the joint program of the U. S. Office of Naval Research and U. S. Atomic Energy Commission.
† Holder of Shell Fellowship 1952–1954.
[1] H. Euler and B. Kockel, Naturwiss. **23**, 346 (1935); W. Heisenberg and H. Euler, Z. Physik **98**, 714 (1936); H. Euler, Ann. Physik. **26**, 398 (1936); V. Weisskopf, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **14**, No. 6 (1936).
[2] M. Delbrück, Z. Physik **84**, 144 (1933); A. Achieser and I. Pomerantschuk, Physik. Z. Sowjetunion **11**, 478 (1937); N. Kemmer, Helv. Phys. Acta **10**, 112 (1937); N. Kemmer and G. Ludwig, Helv. Phys. Acta **10**, 182 (1937); F. Rohrlich and R. Gluckstern, Phys. Rev. **86**, 1 (1952); H. A. Bethe and F. Rohrlich, Phys. Rev. **86**, 10 (1952).
[3] O. Halpern, Phys. Rev. **44**, 855 (1933); H. Euler and B. Kockel, reference 1; H. Euler, reference 1; R. Karplus and M. Neuman, Phys. Rev. **80**, 380 (1950) and **83**, 776 (1951).
[4] E. J. Williams, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **13**, No. 4 (1935).
[5] R. R. Wilson, Phys. Rev. **90**, 720 (1953).

due to a factor $k_1^3k_2^3$ (a continuous range of energies for each of the two scattered photons is allowed by the conservation laws: $\mathbf{k}=\mathbf{k}_1+\mathbf{k}_2+\mathbf{K}$ and $k=k_1+k_2+K^2/2M \approx k_1+k_2$). In the case of double Compton scattering from electrons[6] (a competing process with a cross section per atom very roughly $(2\pi^2)^{-1}\alpha(e^2/mc^2)^2Z=2.4\times10^{-27}$ cm^2/sterad^2), for the same configuration, the conservation equations for the energies of the two product photons have the form $\mathbf{k}=\mathbf{K}$, $\mathbf{k}_1=\mathbf{k}_2$, and $k+1=k_1+k_2+(K^2+1)^{\frac{1}{2}}$. For $k=1.65$, these give $k_1+k_2=0.72$. The energy of a single-Compton scattered photon under these circumstances is 0.62 (random coincidences might compete here because of the relatively huge cross section). The cross section for double Compton scattering from nuclei is relatively negligible (approximately $(2\pi^2)^{-1}Z^2\alpha[(Ze)^2/AMc^2]^2[kmc^2/AMc^2]^2 \approx 10^{-37}$ cm^2/sterad^2). Thus, by biasing counters to accept only photons in coincidence with each photon energy lying between 0.7 and $1.65-0.7=0.95$ and requiring that the two energies add to 1.65, most of the background scattering should be eliminated; moreover, if photons are accepted only in this energy range, the cross section is still about 7×10^{-34} cm^2/sterad^2.

This cross section for the splitting of a photon into two photons has been calculated (in lowest order in α) for small incident momenta ($k<1$) from the equivalent vacuum polarization Hamiltonian of Euler and others:[7]

$$H_I=-(\alpha^2/360\pi^2)[(\hbar/mc)^3/mc^2][(\mathbf{D}^2-\mathbf{B}^2)^2+7(\mathbf{D}\cdot\mathbf{B})^2]$$

of which the contributing terms are (nuclear electrostatic field $\mathbf{D}_n$, photon field $\mathbf{D}_p$, $\mathbf{B}_p$; $\mathbf{D}=\mathbf{D}_p+\mathbf{D}_n$, $\mathbf{B}=\mathbf{B}_p$):

$$-(\alpha^2/180\pi^2)[(\hbar/mc)^3/mc^2]\times[2\mathbf{D}_p^2\mathbf{D}_p\cdot\mathbf{D}_n-2\mathbf{B}_p^2\mathbf{D}_p\cdot\mathbf{D}_n+7\mathbf{D}_p\cdot\mathbf{B}_p\mathbf{D}_n\cdot\mathbf{B}_p].$$

By using a shielded potential for the nucleus,

$$\mathbf{D}_n=-Ze\,\mathrm{grad}\{[\exp(-r/a)]/r\},$$

the differential cross section for splitting of the incident photon of momentum $\mathbf{k}$ into two photons of momenta $(\mathbf{k}_1,\mathbf{k}_1+d\mathbf{k}_1)$ in $d\Omega_1$, $\mathbf{k}_2$ in $d\Omega_2$, is obtained (wavenumbers in units of $mc/\hbar$, a in units of $\hbar/mc$):

$$\sigma dx d\Omega_1 d\Omega_2=\frac{\alpha^5(\hbar/mc)^2Z^2k^6}{32\cdot81\cdot25\cdot\pi^4}\cdot\frac{x^3(1-x)^3(Ax^2+Bx+C)}{[\mathbf{K}^2/2k^2+1/2a^2k^2]^2}dx d\Omega_1 d\Omega_2\ \mathrm{cm}^2,$$

$$\int_0^1\sigma dx=Qk^6Z^2\ \mathrm{cm}^2/\mathrm{sterad}^2,$$

where $\mathbf{K}=\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2$, $k_2=k-k_1$, $x=k_1/k$, and A, B, C are functions of the angles involved, say, $\cos(\mathbf{k},\mathbf{k}_1)=\cos\theta_1=u$; $\cos(\mathbf{k},\mathbf{k}_2)=\cos\theta_2=v$; $\cos(\mathbf{k}_1,\mathbf{k}_2)=\cos\theta_{12}=w$. Then in terms of u, v, and $t=1-w$,

$$\mathbf{K}^2/2k^2=tx^2+x(v-u-t)+1-v,$$

and

$$A=157t^6+t^5(-193+157v+157u-139v^2-139u^2+157uv)+t^4(278-278v-278u+157v^2+157u^2-36uv),$$

$$B=-157t^6+t^5(193-278v-36u+278v^2-157uv)+t^4(-278+363v+193u-278v^2-36u^2+36uv+157uv^2-157u^2v)+(v-u)(278-278v-278u+157v^2+157u^2-36uv),$$

$$C=(1-v)\{t^5(18+139v)+t^4(85-121v-121u+157uv)+139-139v+157v^2+18u^2-314uv+139u^2v\}.$$

For the particular case $\theta_1=\pi/2$, $\theta_2=\pi/2$, $\theta_{12}=\pi$, Q is found to be 1.68×10^{-38} cm^2/sterad^2. The variation of Qk^6Z^2 with, for example, deviations of θ_{12} from π for fixed $\theta_1=\pi/2$, $\theta_2=\pi/2$, is relatively slow.

If $k>10^{-2}$, A, B, and C are very small in the region where the term $1/a^2k^2$ becomes significant (i.e., as $\mathbf{K}^2/2k^2\to0$). Then a rough estimate of the total cross section $\sigma(k)$ is $(4\pi)^2$ times the differential cross section for the configuration above, and works out to be about $0.7\times10^{-2}\alpha^5(\hbar/mc)^2Z^2k^6$ cm^2.

For high-energy incident photons ($k\gg1$), a simple calculation by the Weizsäcker-Williams method,[8] using the cross section for high-energy scattering of photons by photons derived by Achieser,[9] shows that

$$\sigma(k)=b\alpha^5(\hbar/mc)^2Z^2\log(\eta k)\ \mathrm{cm}^2,$$

where η is of the order of 1; a similar result using a cruder estimate of the cross section for high-energy photon-photon scattering [$\sim\alpha^4(\hbar/mc)^2$ instead of $\sim\alpha^4(\hbar/mc)^2(1/k_1k_2)$] has been given by Williams.[4] If the expressions for small k and for large k are equated near $k=2$, then $b\approx1/2$.

The author wishes to thank Dr. H. Primakoff for suggesting this problem and for many helpful discussions.

[6] F. Mandl and T. Skyrme, Proc. Roy. Soc. (London) **215**, 497 (1952).

[7] H. Euler, reference 1; V. Weisskopf, reference 1; J. Schwinger, Phys. Rev. **82**, 664 (1951).

[8] C. V. Weizsäcker, Z. Physik **88**, 612 (1934); E. J. Williams, reference 4.

[9] A. Achieser, Physik Z. Sowjetunion **11**, 263 (1937).

Radioisotope P^{33}†

R. T. NICHOLS AND E. N. JENSEN
Institute for Atomic Research, Iowa State College, Ames, Iowa

(Received August 13, 1953)

Radioactive phosphorus has been obtained by an n, p reaction on sulfur which had been electromagnetically enriched in S^{33}. The beta spectrum of this phosphorus showed an increase in the low energy beta group approximately proportional to the isotopic enrichment. This confirms the previous identification of this activity with P^{33}. The maximum beta energy was determined from several Kurie plots and found to be 0.249 ± 0.002 Mev. The half-life was found to be 24.4 ± 0.2 days. The spectrum appears to have an allowed shape as is expected from theoretical considerations.

I. INTRODUCTION

A LOW-ENERGY beta group has been reported in P^{32} samples[1,2] and identified with P^{33}. For a further study of this activity, radioactive phosphorus was obtained from a neutron irradiation of sulfur which had been electromagnetically enriched in S^{33}. An increase in yield of the low-energy activity approximately proportional to the enrichment of S^{33} was interpreted as confirmation of the isotopic assignment. The increased amount of P^{32} also served to facilitate more accurate measurement of half-life and energy characteristics.

II. EXPERIMENTAL RESULTS

A half gram of sodium sulfate which had been enriched in S^{33} by a mass separation was obtained from Oak Ridge. This sample, along with a control sample of ordinary sodium sulfate, was irradiated in the Chalk River pile for 128 hours. Carrier-free phosphorus sources were prepared from these samples by use of an exchange column.[3]

The initial spectra from these sources were obtained with a thin-lens spectrometer[3] which had been modified to incorporate ring focusing.[4] Figure 1 shows a comparison of the beta spectra of the control and the enriched phosphorus about 3½ weeks after removal from the pile. The spectra have been normalized at the peak region of the P^{32} spectra. The ordinate values are for the spectrum of the enriched phosphorus. The increased yield of the low-energy beta group supports the previous conclusion[2] that it is the activity of P^{33}.

Additional spectra were obtained periodically of the P^{33}-enriched phosphorus source for half-life and end-point energy determinations of P^{33}. When the source had become too weak for the thin-lens spectrometer, an intermediate image spectrometer[5] was used to extend the investigation. This increased the counting rate approximately 150 times due to a fiftyfold increase in transmission and an increase in half-width of about three.

Kurie plots of the P^{33} component obtained from earlier samples[2] did not appear to be linear, but low statistics and thick source effects precluded any definite conclusions about the experimental spectral shape. The first Kurie plot from the P^{33}-enriched source is shown in Fig. 2 as it appeared 68 days after irradiation. A least squares calculation of the Kurie line yielded a value of 0.248 Mev for the maximum beta energy. The straight line indicates an allowed spectral shape. However, the Kurie plots from later spectra were increasingly nonlinear with the line convex toward the energy axis, as shown by the crosses in Fig. 3. This changing shape suggested the presence of another longer-lived activity of slightly lower energy, perhaps the activity previously reported[2] in half-life data from an old phosphorus source. This activity had a half-life of about 90 days.

A spectrum was taken 279 days after the neutron irradiation. At this time the P^{32} activity was negligible in the energy range of P^{33}. The spectrum was definitely distorted in such a way as to indicate the presence of another beta group of slightly lower energy than that of P^{33}. This may be seen in the total spectrum of Fig. 4. In this case the Kurie plot appeared linear above 175 kev. A least squares line through these points gave a maximum energy value of 0.250 Mev. An S^{35} Kurie

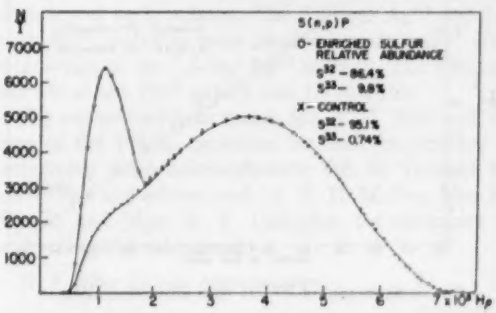

FIG. 1. Beta spectra of phosphorus samples 20 days after irradiation.

† Contribution No. 303 from the Institute for Atomic Research and Department of Physics, Iowa State College, Ames, Iowa. Work was performed in the Ames Laboratory of the U. S. Atomic Energy Commission.

[1] Sheline, Holtzman, and Fan, Phys. Rev. **83**, 919 (1951).

[2] Jensen, Nichols, Clement, and Pohm, Phys. Rev. **85**, 112 (1952).

[3] Jensen, Laslett, and Pratt, Phys. Rev. **75**, 458 (1949).

[4] Pratt, Boley, and Nichols, Rev. Sci. Instr. **22**, 92 (1951); Keller, Koenigsberg, and Paskin, Rev. Sci. Instr. **21**, 713 (1950).

[5] Nichols, Pohm, Talboy, and Jensen (to be published).

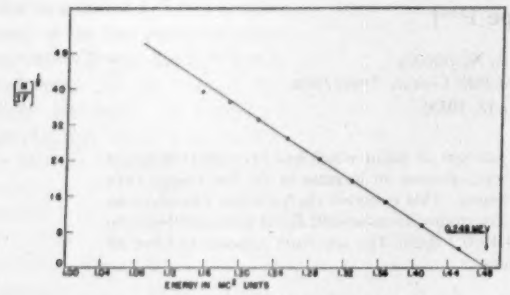

FIG. 2. Kurie plot of P^{32} 68 days after irradiation.

plot of the remainder of the spectrum gave an end-point energy of 0.170 Mev which agrees reasonably well with previous values[6] for S^{35}. These two Kurie plots are shown in Fig. 5. The half-life of the activity remaining in the sample 416 days after removal from the pile was found to be about 85 days. It is estimated that the activity due to P^{33} was less than three percent of the total activity. The components of the total beta spectrum are indicated in Fig. 4.

Earlier spectra from the P^{33}-enriched source were corrected for S^{35} activity by calculating the relative change of these components and normalizing the calculated S^{35} activity by matching the P^{33} components in the region above the maximum energy of S^{35}. A value of 24.4 days was used for the half-life of P^{33} and a value of 87.5 days was used for the half-life of S^{35}. After the calculated S^{35} activity was subtracted, the Kurie plots were linear as shown by the circles in Fig. 3.

The relative growth of P^{33} to P^{32} over a period of 191 days was determined from five spectra. The area of the P^{33} and P^{32} spectrum components was compared in each spectrum using only the region above 120 kev to insure no change in transmission efficiency due to window absorption in the Geiger counters or pulse discrimination in the scintillation counter. A logarithmic plot of the relative activities gave an excellent straight line

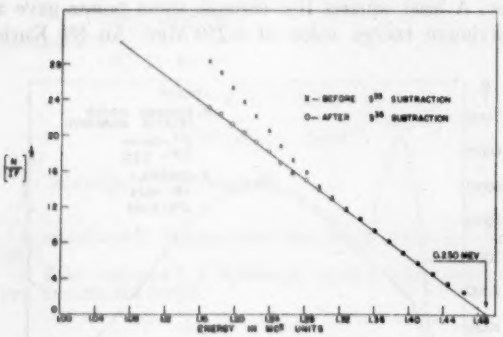

FIG. 3. Kurie plot of P^{33} 211 days after irradiation.

[6] Hollander, Perlman, and Seaborg, Revs. Modern Phys. **25**, 469 (1953).

and yielded a half-life value of 24.3 days for P^{33} using a value of 14.3 days for the half-life of P^{32}.

By extension of the Kurie plots of P^{32} and P^{33}, the total ratio of the activities was found for one spectrum and calculated back to the time of removal of the sample from the pile. The relative cross section for the n, p reaction on S^{33} compared to that of S^{32} for the neutron distribution in the Chalk River pile was found to be about 2.6. This increase may be expected since the $S^{32}(n,p)P^{32}$ reaction is endoergic, requiring neutron energies of about 1 Mev, while the $S^{33}(n,p)P^{33}$ reaction is exoergic and may be induced by slower neutrons.

III. SUMMARY

The results of this investigation confirm the previous conclusion that the low energy beta group in P^{32} samples formed by an n, p reaction on sulfur is due to P^{33}. With revaluation of previous half-life data in light of

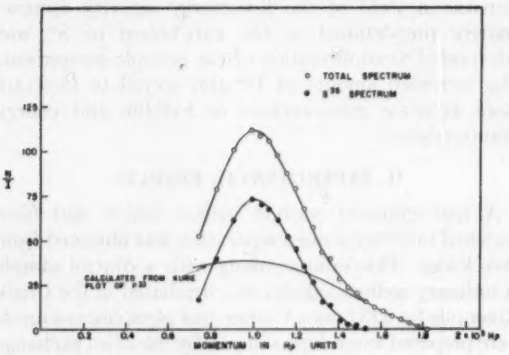

FIG. 4. Beta spectrum of phosphorus from S^{32}-enriched sulfur 279 days after irradiation.

the observed S^{35} in the phosphorus samples, the mean value of the half-life of P^{33} is now found to be 24.4 ± 0.2 days. The maximum beta energy, taken from four linear Kurie plots and both spectrometers, is calculated to be 0.249 ± 0.002 Mev.

Using these values, an f value of 0.055 is obtained from the graphs of Feenberg and Trigg[7] and a log ft value of 5.1 is found for P^{33}. According to Nordheim[8] this is a normal allowed transition. The spin of S^{33} is known to be $\frac{3}{2}$.[9,10] The magnetic moment of S^{33} indicates a $d_{\frac{3}{2}}$ state.[10] According to the nuclear shell model,[11] a regular filling of the levels would give a $d_{\frac{3}{2}}$ state to the P^{33} nucleus. The decay of P^{32} would then be a normal allowed transition in which the orbital momentum and spin are unchanged. For such a transition the values of

[7] E. Feenberg and G. Trigg, Revs. Modern Phys. **22**, 399 (1950).
[8] L. W. Nordheim, Phys. Rev. **78**, 294 (1950).
[9] C. H. Townes and S. Geschwind, Phys. Rev. **74**, 626 (1948).
[10] Eshbach, Hillger, and Jen, Phys. Rev. **80**, 1106 (1950).
[11] M. G. Mayer, Phys. Rev. **78**, 16 (1950).

log *ft* range mostly from 4.8 to 5.5.[8] The experimental value of 5.1 thus agrees well with these values and the experimental spectral shape is consistent with these considerations. However, P^{31} is known to have a spin of $\frac{1}{2}$ and an $s_{\frac{1}{2}}$ state. On the basis of the nuclear shell model one would expect P^{33} to have an $s_{\frac{1}{2}}$ state also. The decay of P^{33} would then involve an orbital momentum change of two and a spin change of one. This type of transition has log *ft* values mostly in the range of 6.5 to 7.5,[8] but with some stragglers. The experimental value of 5.1 for log *ft* is not in agreement with these values. These considerations add confirmation to the $d_{\frac{1}{2}}$ state for P^{33}.

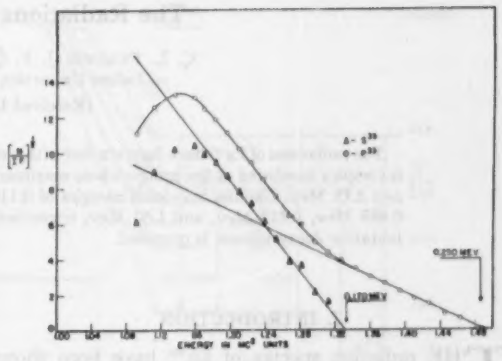

FIG. 5. Kurie plots of P^{33} and S^{35} 279 days after irradiation.

The authors wish to express their appreciation to Dr. A. F. Voigt and Mr. L. McIsaac for chemical separation of the phosphorus from irradiated sulfur and for preparation of the spectrometer source. Thanks are also due Mr. W. E. Lewis, Mr. J. T. Jones, Jr., and Mr. E. W. McMurry for assistance in obtaining the data.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

A New Isotope of Palladium, 1.5-Minute Pd^{113}†

HARRY G. HICKS AND RICHARD S. GILBERT
California Research and Development Company, Livermore, California

(Received October 9, 1953)

A new isotope, 1.5-minute Pd^{113}, has been isolated from the fission products of natural uranium bombarded with 190-Mev deuterons. The mass assignment and half-life were determined by successive milkings of the 5.3-hour Ag^{113} daughter.

EXAMINATION of the neutron excess isotopes of palladium led the authors to speculate that Pd^{113} may have a half-life which could be determined by milking the well-known 5.3-hour Ag^{113} daughter.

A strip of uranium foil was bombarded with 190-Mev deuterons for two minutes in the circulating beam of the Berkeley 184-in. cyclotron using a "rabbit-type" target. After bombardment the target was transported in a pneumatic tube to the chemistry laboratory and dissolved in aqua regia with palladium holdback agent added. Two silver chloride scavenges were performed, the second scavenge was complete 4.5 minutes after the beam was turned off. The supernatant liquid from the second scavenge was decanted into a centrifuge cone containing a measured amount of silver nitrate solution. The mixture was maintained at about 90°C and stirred vigorously. The precipitate was centrifuged so that the time of separation of the precipitate from the solution was, as nearly as possible, one minute after the separation of the silver chloride scavenge. Twelve milkings were made at one minute intervals. The silver chloride precipitate was washed as soon as possible after decantation. Less than two percent of the supernatant liquid remained with the decanted precipitate.

After all the milkings and washings were complete, silver was purified radiochemically to remove radionuclides that had adsorbed on the silver chloride. The purified samples were counted with a standard chlorine-quenched Geiger tube. The 7.5-day Ag^{111} activity milked from the 22-minute Pd^{111} was resolved from the decay of the silver samples and used to determine the efficiency of each milking. The 3.2-hour Ag^{112} and 5.3-hour Ag^{113} activities were resolved analytically[1] after subtraction of the 7.5-day Ag^{111} activity. The resultant half-life of the Pd^{113} parent was 1.5 minutes.

The authors wish to thank Mr. J. T. Vale and the crew of the 184-in. cyclotron for their cooperation in performing these bombardments, Dr. M. Lindner for his interest and advice, and Dr. W. H. McVey, Mrs. M. Nervik, and Miss M. F. Gallagher for assistance in performing the experiments.

† This work was done under the auspices of the U. S. Atomic Energy Commission.

[1] W. F. Biller (private communication).

The Radiations of Lanthanum-140

C. L. Peacock, J. F. Quinn, and A. W. Oser, Jr.
Tulane University, New Orleans, Louisiana
(Received December 7, 1953)

The radiations of La^{140} have been studied with a magnetic spectrometer of radius 20 cm. The beta spectrum is complex consisting of five groups whose maximum energies are: 0.83 Mev, 1.10 Mev, 1.34 Mev, 1.67 Mev, and 2.15 Mev. Gamma rays with energies of 0.110 Mev, 0.130 Mev, 0.240 Mev, 0.270 Mev, 0.328 Mev, 0.485 Mev, 0.815 Mev, and 1.60 Mev, respectively, were found. The 1.60-Mev gamma is very intense. A tentative decay scheme is proposed.

I. INTRODUCTION

THE radiation spectra of La^{140} have been shown by previous studies to be quite complex.[1] The results of these studies were summarized in various proposed decay schemes. A later report[2] on the gamma radiations of Ce^{140} following the beta transitions of La^{140} noted the presence of fifteen gamma-ray energies with varying intensities. A decay scheme which differed from all previous ones was proposed on the basis of the energies of these gamma radiations.

The present work was undertaken as a further study of this isotope in an attempt to formulate its decay scheme. Although particular attention was directed toward the beta-ray energies, the gamma rays were also studied.

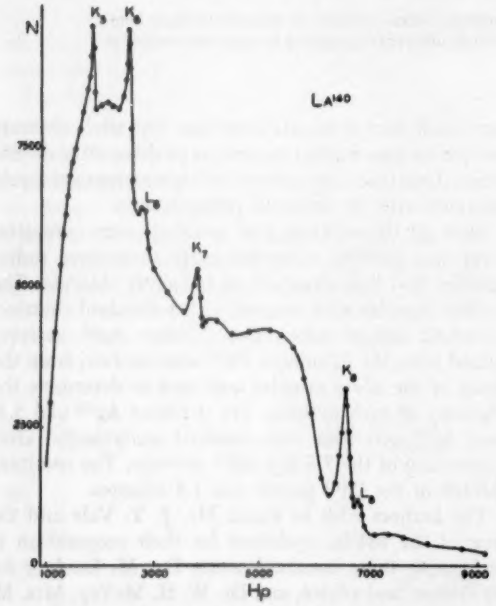

FIG. 1. The photoelectron spectrum of the gamma rays of La^{140} from a lead radiator.

[1] A summary of references to studies made before 1949 can be found in the article by Beach, Peacock, and Wilkinson, Phys. Rev. 76, 1624 (1949).
[2] J. M. Cork *et al.*, Phys. Rev. 83, 856 (1951).

APPARATUS AND METHODS

This study was carried out with a 180-degree type of magnetic spectrometer of radius 20 cm. The magnet was of the ring-shaped type. It was used to give a homogeneous field inside the vacuum chamber. Both conventional mica window and very thin Zapon film window G-M tubes were used as detectors. The resolution of the instrument was set at about 2 percent.

Sources were prepared from lanthanum which had been chemically separated from an equilibrium mixture of barium[140] and lanthanum[140]. The source material was obtained from the Oak Ridge National Laboratory. The lanthanum was precipitated as $La(OH)_3$ by the addition of NH_4OH to the equilibrium mixture. It was then converted to the chloride by the addition of HCl. Repeated precipitations assured a barium free sample. The beta sources, which were evaporated on very thin films of LC 600, had a thickness of less than 0.1 mg/cm².

TABLE I. Gamma radiations of La^{140}.

K_1	0.110 Mev	K_5	0.328 Mev
K_2	0.130 Mev	K_6	0.485 Mev
K_3	0.240 Mev	K_7	0.815 Mev
K_4	0.270 Mev	K_8	1.60 Mev

III. RESULTS

Figure 1 shows the spectrum of photoelectrons ejected from a lead radiator as detected by a mica window counter. The photoelectron peaks observed in this distribution can be interpreted to correspond to gamma rays of the following energies: 0.328 Mev, 0.485 Mev, 0.815 Mev, and 1.60 Mev. (The numbering system in the figure corresponds to the designations in Table I.)

Figure 2 shows a fitted composite of several studies made on the electron spectrum. Various portions of the curve were given particularly careful study at different times with the aim of locating the lines due to the internally converted gamma rays already reported[2] as well as searching for others. These portions were then corrected for decay and fitted together. The lines observed in this curve may be attributed to gamma rays whose energies are as follows: 0.110 Mev, 0.130 Mev, 0.240 Mev, 0.270 Mev, 0.328 Mev, 0.485 Mev, 0.815

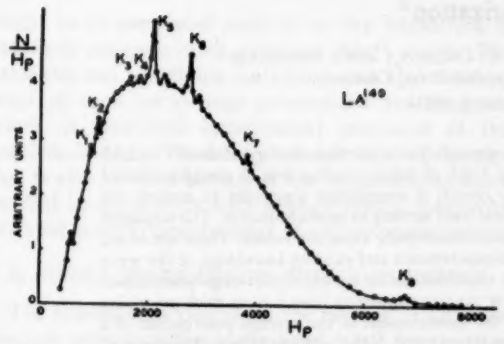

FIG. 2. An electron spectrum of La^{140}.

Mev, and 1.60 Mev. It is to be noted that those peaks found in the photoelectron spectrum are also contained in this latter group. A summary of the results of the gamma-ray studies is presented in Table I. The intensities of the first four gamma rays could not be determined from this work due to the small magnitude and the complicated decay scheme. The ratio of the intensities of the last four gamma rays as determined

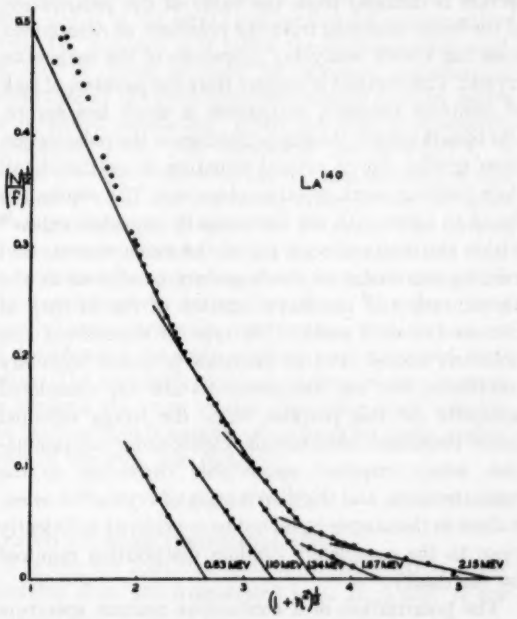

FIG. 3. A Fermi analysis of the beta rays of La^{140}.

from the photoelectric conversion is 0.328:0.485:0.815: 1.60 as 1:1:1:10. However, the 0.328-Mev and 0.485-Mev gamma rays show a greater internal conversion than do the others. It is noted that these energies compare favorably with reported values.[3] Some of the reported lines, however, were not found in this study.

The distribution of beta rays was analyzed by the Fermi plot method with the aid of tables prepared at

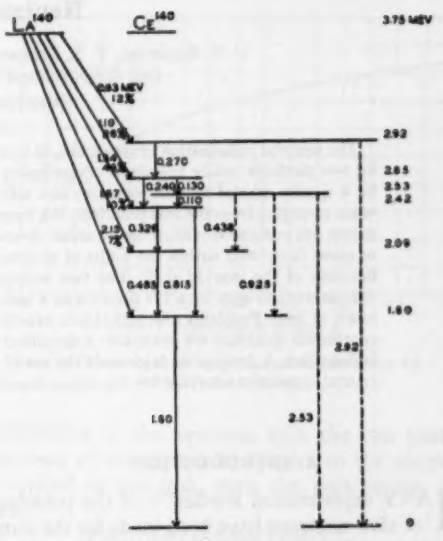

FIG. 4. Suggested disintegration scheme of La^{140}.

the Bureau of Standards.[3] Those points corresponding to electron lines due to converted gamma rays have been omitted from the analysis. Figure 3 shows a typical plot of a representative beta-ray distribution. It is noted that this plot indicates complexity. In fact, the plot seems to have definite curvature in the region below 1.34 Mev. This could denote forbiddenness of the 1.34-Mev group of beta rays. If, however, the groups are assumed to have straight line plots, the distribution plot can be analyzed into groups whose maximum energies are: 2.15 Mev, 1.67 Mev, 1.34 Mev, 1.10 Mev, and 0.83 Mev. It is felt that the deviation below 0.4 Mev is due mainly to source thickness.

From the values of the gamma- and beta-ray energies obtained in this study, a disintegration scheme can be formulated. This scheme is shown in Fig. 4. The 1.60-Mev gamma ray has been found to be in coincidence with the highest energy beta-ray group, as well as with the 0.815-Mev gamma ray.[4,5] Other gamma-ray energies fit remarkably well between the energy levels suggested by the beta-ray analysis. These are denoted in the decay scheme by the solid lines. Two of the gamma energies reported by Cork *et al.*, but not found in this study, can also be fitted into the scheme, as well as the 2.53 Mev gamma-ray energy reported by Wattenberg[6] and the 2.92 Mev one reported by Bishop *et al.*[7] These are denoted by the dotted lines in Fig. 4.

---

[3] L. F. Curtiss and I. Feister, *Tables for the Analysis of Beta Spectra*, National Bureau of Standards, Applied Mathematics Series No. 13 (U. S. Government Printing Office, Washington, D. C., 1951).

[4] B. L. Robinson and L. Madansky (to be published).

[5] R. G. Wilkinson (private communication).

[6] A. Wattenberg, Phys. Rev. **71**, 497 (1947).

[7] Bishop, Wilson, and Halban, Phys. Rev. **77**, 416 (1950).

Neutron Polarization*

C. P. STANFORD, T. E. STEPHENSON, L. W. COCHRAN,† AND S. BERNSTEIN
Oak Ridge National Laboratory, Oak Ridge, Tennessee
(Received November 9, 1953)

The neutron polarization cross section of iron has been measured as a function of energy from 0.7 to 3.3A by two methods: using the single transmission effect in a block of polycrystalline iron at energies selected by a quartz crystal monochromator; and using a single crystal of magnetized magnetite to analyze the beam emerging from the iron polarizer, the magnetite crystal itself serving as monochromator. The measured values are compared with those of other observers and the theoretically expected values. These are found to agree fairly well within the limits of accuracy of the measurements and existing knowledge of the wave function of the iron *3d* shell. The two techniques were used also to determine the average polarization (32 percent) as seen by a $1/v$ detector in a beam of reactor neutrons emerging from a 4-cm thick polarizing block of iron. Problems and techniques associated with the measurement of the average polarization of a continuous spectrum are discussed. A simplified experimental treatment of the problem of beam "hardening" is described. A description is given of the use of the magnetic resonance method in conjunction with a single-crystal magnetite analyzer for the measurement of neutron polarization.

A. INTRODUCTION

MANY experimental studies[1–12] of the polarization of slow neutrons have been made for the purpose of verifying the theory[13–21] of scattering, polarization, and depolarization of neutrons passing through crystalline magnetic materials. Some of the experiments measure the energy dependence of the change in the total cross section upon magnetization of a polycrystalline sample of iron. The rotating shutter[8] and the pulsed cyclotron[9] time of flight methods have been used for the velocity selection. We have used the higher resolution of the single-crystal spectrometer to measure the polarization cross section as a function of energy from about 0.7 to 3.3A, by two methods: using the single transmission effect in a block of polycrystalline iron at energies selected by a quartz crystal monochromator; and using a single crystal of magnetized magnetite to analyze the beam emerging from the iron polarizer, the magnetite crystal itself serving as monochromator. In the first of these methods the polarization cross section is deduced from the change in the intensity of the beam transmitted by the polarizer on magnetization. We have developed the application of the second method, in which the polarization cross section is deduced from the value of the polarization of the beam emerging from the polarizer, as determined from the known analyzing properties of the magnetite crystal. This method is simpler since the problem of lack of complete magnetic saturation is much less severe. Our results exhibit the discontinuities in the polarization cross section due to crystal structure in greater detail than previous work of other observers. The results are found to agree with the theoretically expected values[20] within the limits of accuracy of the measurements and existing knowledge of the boundary conditions at the atomic radius of the wave function of the *3d* shell of iron in the solid state. This type of experiment can definitely choose between alternate proposed boundary conditions, but our measurements are not considered adequate for this purpose, since the Bragg reflected beam contained considerable higher-order contamination which required appreciable correction to the measurements, and the distribution of crystallite orientations in the sample could not be considered sufficiently close to the completely random distribution required by the theory.

The polarization of a continuous neutron spectrum is difficult to measure with accuracy by use of the single transmission effect because the polarizing block alters the spectrum of the neutron beam. Also, the variation of detector sensitivity with energy may complicate the interpretation of the intensity changes. The two methods described above were used to determine the average polarization (32 percent) seen by a $1/v$ detector in a beam of reactor neutrons emerging from a 4-cm thick polarizing block of iron. This information is per-

* Based, in part, on a dissertation being submitted by C. P. Stanford in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Tennessee.

† Oak Ridge National Laboratory, Research Participant, Department of Physics, University of Kentucky, Louisville, Kentucky.

[1] Hoffman, Livingston, and Bethe, Phys. Rev. **51**, 214 (1937).
[2] Frisch, Halban, and Koch, Phys. Rev. **53**, 719 (1938).
[3] Powers, Carroll, and Dunning, Phys. Rev. **51**, 1112 (1937).
[4] Dunning, Powers, and Beyer, Phys. Rev. **51**, 51 (1937).
[5] Bloch, Hamermesh, and Staub, Phys. Rev. **64**, 47 (1943).
[6] E. M. Fryer, Phys. Rev. **70**, 235 (1946).
[7] Bloch, Condit, and Staub, Phys. Rev. **70**, 972 (1946).
[8] Hughes, Wallace, and Holtzman, Phys. Rev. **73**, 1277 (1948).
[9] Fleeman, Nicodemus, and Staub, Phys. Rev. **76**, 1774 (1949).
[10] Burgy, Hughes, Wallace, Heller, and Woolf, Phys. Rev. **80**, 953 (1950).
[11] D. J. Hughes and M. Burgy, Phys. Rev. **81**, 498 (1951).
[12] Shull, Wollan, and Koehler, Phys. Rev. **84**, 912 (1951).
[13] F. Bloch, Phys. Rev. **50**, 259 (1936); **51**, 994 (1937).
[14] J. Schwinger, Phys. Rev. **51**, 544 (1937).
[15] O. Halpern and M. H. Johnson, Phys. Rev. **55**, 898 (1939).
[16] O. Halpern and T. Holstein, Phys. Rev. **59**, 960 (1941).
[17] Halpern, Hamermesh, and Johnson, Phys. Rev. **59**, 981 (1941).
[18] M. Hamermesh, Phys. Rev. **61**, 17 (1942).
[19] O. Halpern, Phys. Rev. **76**, 1130 (1949).
[20] J. Steinberger and G. C. Wick, Phys. Rev. **76**, 994 (1949).
[21] H. Ekstein, Phys. Rev. **76**, 1328 (1949).

tinent to an associated project[22] on the interaction of polarized neutrons with polarized Mn^{55} nuclei. The problems and techniques associated with the measurement of such an average polarization are discussed below. A simplified experimental treatment of the problem of beam "hardening" is described. A description is also given of the use of the magnetic resonance method in conjunction with a single-crystal magnetite analyzer for the measurement of neutron polarization.

B. SINGLE TRANSMISSION EFFECT *VS* ENERGY

The relationships describing the passage of neutrons through polycrystalline magnetized iron[16] for the case in which the incident beam is unpolarized are given by

$$j=j_0e^{-\frac{1}{2}qd}[\cosh(d/\Delta)+\tfrac{1}{2}q\Delta\sinh(d/\Delta)], \quad (1)$$

$$S=j_0e^{-\frac{1}{2}qd}w\Delta\sinh(d/\Delta), \quad (2)$$

where j is the intensity of neutrons emerging from the magnetized iron block and j_0 is the intensity emerging from the iron block when it is unmagnetized. j_0 is equal to the intensity incident upon the block times the factor $e^{-N\sigma_i d}$, in which N is the number of scatterers per unit volume in the iron, d is the thickness of the iron, and σ_i is the total cross section for unmagnetized iron. The quantity q is called the depolarization coefficient. It takes account of relevant metallurgical properties of the polycrystalline sample and the degree of magnetic saturation. When the iron is completely saturated, $q=0$. The parameter $\Delta=(w^2+\frac{1}{4}q^2)^{-\frac{1}{2}}$, in which $w=Np$, is the polarization cross section per unit volume of iron. The scattering cross section per atom of magnetized iron is given by $\sigma=\sigma_0\pm p$. The term σ_0 is the scattering cross section for unmagnetized iron. It includes the scattering due to nuclear interaction alone and magnetic interaction alone. The quantity p is called the polarization cross section. It arises because of interference between nuclear and magnetic scattering. The scattering due to interference may increase or decrease the cross section depending upon whether the neutron spin is parallel or antiparallel to the atomic spins of the iron. The quantity

$$(j/j_0)-1=E_1 \quad (3)$$

is called the single transmission effect.

S is the net spin flux density of the neutron beam emerging from the magnetized iron. If j_1 and j_2 are the intensities of the neutrons with the two possible orientations of the spins with respect to the magnetic field applied to the iron, then the polarization P is given by

$$P=S/j=(j_1-j_2)/(j_1+j_2). \quad (4)$$

In the case of complete saturation the relation between the polarization and the intensities can be expressed by

$$P=[1-(j_0/j)^2]^{\frac{1}{2}}. \quad (5)$$

This relationship is plotted in Fig. 1.

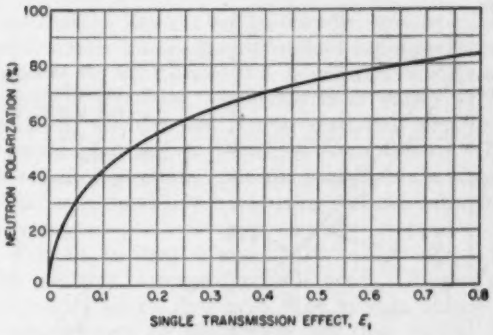

FIG. 1. Polarization *vs* single transmission effect for complete saturation.

For a single neutron energy, in the absence of complete saturation, measurements of E_1 for two different thicknesses of iron in the same state of magnetization will, in principle, determine the values of q and Δ. From Δ, the value of p, the polarization cross section can be derived. Using these values of q and Δ in (1), (2), and (4), the polarization can be determined. The quantity w, and therefore Δ also, depend upon energy. We have used the method just outlined to determine q and Δ for neutrons of a single energy. The grain size of the iron samples and the neutron spectrum were such that q was taken to be independent of energy. The measured value of Δ was used to fix the scale on the known shape of the curve of polarization cross section *vs* energy. The polarization as a function of energy was then calculated from (1), (2), and (4). The polarization averaged over the spectrum was calculated by weighting the polarization for a given energy in proportion to the abundance of neutrons of that energy in the beam emerging from the polarizing block of iron.

The neutron energy 0.0372 ev was selected by Bragg reflection from a single crystal of quartz following the polarizing block of iron. The single transmission effect was measured for 1, 2, 3, and 4 cm of iron. Care was taken to demagnetize the iron thoroughly before the field-off count. The arrangement was such that small angle scattering effects[23] were not likely to influence the results. The state of magnetization of the blocks was taken to be the same when the magnetic field at the

TABLE I. Single transmission effect E_1 for 0.0372 ev.

Thickness of block, cm	E_1, measured	E_1, corrected
1	0.020±0.001	0.020
2	0.071±0.001	0.074
3	0.147±0.003	0.157
4	0.253±0.003	0.274

[22] Bernstein, Roberts, Stanford, Dabbs, and Stephenson, Phys. Rev. (to be published).

[23] Hughes, Burgy, Heller, and Wallace, Phys. Rev. 75, 565 (1949).

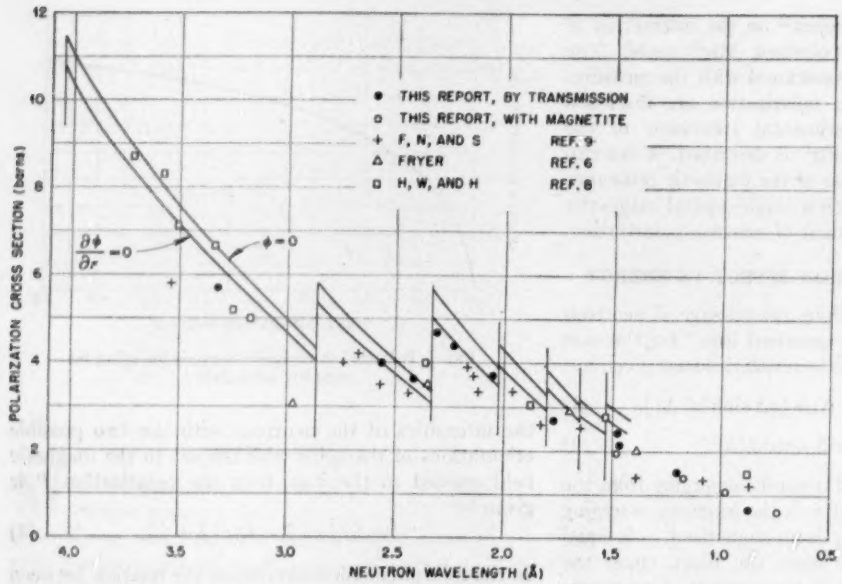

FIG. 2. Polarization cross section *vs* energy. The points give the measured values of this report and those of other observers. The curves are calculated from (7) using σ_{coh} for iron = 11.4 barns.

surface of each of the blocks was measured to be the same with a small flat flip coil which could be placed up against the surface of the block. The field at the surface of each block was 11 000 oersteds. The results of these measurements are given in Table I.

The direct measurements must be corrected for the presence of neutrons of energy 4×0.0372 ev due to the second-order Bragg reflection from the quartz crystal. Orders of reflection higher than the second present in the reflected beam were estimated to have negligible effect upon the measurements because the intensity of neutrons incident upon the crystal is much less for third order, and also because of the decreased reflectivity of the crystal for the higher energy. Corrections for the second order were made by measuring the transmission of a calibrated boron absorber in the beam emerging from each of the blocks. The absorber was calibrated by measuring its transmission at several energies with a mechanical velocity selector. These values were checked by measuring the transmission, also, with a crystal spectrometer arrangement using a single energy so chosen that the second order could be eliminated by means of an appropriate neutron resonance filter. The transmission of the absorber could then be calculated at any energy from the known $1/v$ variation of the boron cross section. The single transmission effect corrected to include first-order neutrons only is given in Table I. The corrected data were fitted using Eqs. (1) and (3) for assumed values of q and w. The values $q=0.204$, $w=0.204$ were found to fit the observations best.

Values of p were measured by the single transmission effect method for several other energies also. In addition, values of p for several energies were derived by a method described below in which the polarization of a single energy in the beam emerging from the iron was measured using a single crystal of magnetite as analyzer. Our measured values of p are shown along with those of other observers in Fig. 2. The polarization cross section is given by the expression[30]

$$p=(\sigma_{coh}/4\pi)^{\frac{1}{2}}(e^2/mc^2)\mu_n\mu_e(\lambda/2a)^2\sum_{l<2a/\lambda}N(l)/l \times(1+\lambda^2l^2/4a^2)F(l), \quad (6)$$

in which σ_{coh} is the coherent nuclear scattering cross section of iron, e, m, and c have their usual meanings, μ_n is the magnetic moment of the neutron in nuclear Bohr magnetons, μ_e is the magnetic moment per iron atom in atomic Bohr magnetons, λ is the neutron wavelength, and a is the dimension of the body-centered-cubic iron unit cell. The quantity $l^2=n_1^2+n_2^2+n_3^2$, where n_1, n_2, and n_3 are the Miller indices of a set of planes. $N(l)$ is twice the value of the usual multiplicity of the set of planes identified by l. $F(l)$ is the magnetic form factor of the iron for the set of planes l. The value of F depends upon the $3d$ electron wave functions of the iron. Values of p have been calculated[30] as a function of energy using the value 10.0 barns for the coherent scattering cross section of iron, on the basis of two assumptions: (1) the wave function ϕ of the $3d$ shell is zero at the atomic radius; and (2) the derivative of the wave function is zero at the atomic radius. We have recalculated p *vs* energy using the more recent value[34] of 11.4 barns for the coherent scattering cross section of iron. The curves are shown in Fig. 2 for comparison with the measured values.

[34] C. G. Shull and E. O. Wollan, Phys. Rev. **81**. 527 (1951).

The measured values fall fairly close to the calculated curve. The accuracy of the measurements is, however, not sufficient to choose between the two assumptions for the magnetic form factor. Equation (6) is based upon the assumption of a random distribution of the orientation of the many small crystallites in the "powder" sample. In the case of our measurements, we used as polarizer, a block of cold-rolled steel, with the magnetic field applied in the direction of rolling. Such an arrangement was chosen because the preferred orientation of the crystallites gives an enhanced effect for a given strength of applied field.[8] It is remarkable that in spite of the known departure of our sample from random orientation, the measured values fall as close as they do to the theoretical curve.

C. TEMPERATURE OF THE INCIDENT NEUTRON BEAM

The assumption was made that the spectrum of neutrons emerging from the reactor from energy zero up to the energy of the Cd cutoff could be represented to a degree of accuracy sufficient for our purposes by a Maxwell distribution. This assumption is known to be only approximately true since the actual spectrum of neutrons emerging from the reactor has fewer neutrons at the very low energies and more neutrons in the 0.5-ev region than does a Maxwell distribution.

We studied the "hardening" of the neutron beam on passage through 1, 2, 3, 4, and 7 cm of iron by measuring the transmission of the calibrated boron absorber in the beam emerging from the iron using a $1/v$ detector. The values of the measured transmissions are given in Table II. The transmission of the absorber rises at first as the thickness of iron is increased, but then falls as thickness is further increased. The measured values were compared to those calculated for the emerging beam, whose spectrum was taken to be the incident Maxwell distribution times the factor $e^{-N\sigma t}$. All integrations were done numerically. The calculated transmissions corresponding to a temperature of 495°K, which were found to come acceptably close to the measured values, are shown in Table II.

The total cross section of iron has not been measured in any great detail for low velocities. In the calculations we have used the curve given by the U. S. Atomic Energy Commission Neutron Cross Section Group[25] as a basis for capture, isotope-disorder, spin-dependent, and inelastic scattering contributions and superimposed upon it the fluctuations due to coherent nuclear and magnetic scattering, calculated from values of the magnetic and coherent nuclear scattering amplitudes furnished by neutron diffraction measurements.[13] The effect of temperature upon the heights of the discontinuities in the coherent scattering was also taken into account.

TABLE II. Transmission of boron absorber.

Thickness of iron cm	Measured transmission	Calculated transmission
1	0.370±0.0004	0.370
2	0.377±0.0003	0.374
3	0.372±0.0006	0.3745
4	0.365±0.0009	0.373
7	0.328±0.004	0.300

[25] *Neutron Cross Sections*," Atomic Energy Commission Report AECU-2040 (Technical Information Service, U. S. Department of Commerce, 1952).

For an incident Maxwell distribution of 495°K, the spectrum corresponding to filtration through 4 cm of iron is shown in Fig. 3. The average energy of those neutrons of wavelength less than 4.04A, the coherent-scattering cutoff, is increased by passage through the iron, causing a "hardening" of the beam. The filtering action of the iron has the effect also of increasing the relative importance of those neutrons of wavelength greater than 4.04A. For 4 cm of iron, they constitute 8 percent of the neutrons in the emerging beam. The variation in relative importance of these two groups of neutrons as a function of thickness explains the presence of the maximum in the boron absorber transmission data.

D. SINGLE TRANSMISSION EFFECT FOR THE CONTINUOUS SPECTRUM

In order to determine the average polarization seen by a $1/v$ detector in a beam of reactor neutrons emerging from 4 cm of polarizing iron, the single transmission effect for all energies less than the Cd cutoff was measured, using a thin $B^{10}F_3$ proportional counter detector with 1, 2, 3, 4 cm of iron between the poles of

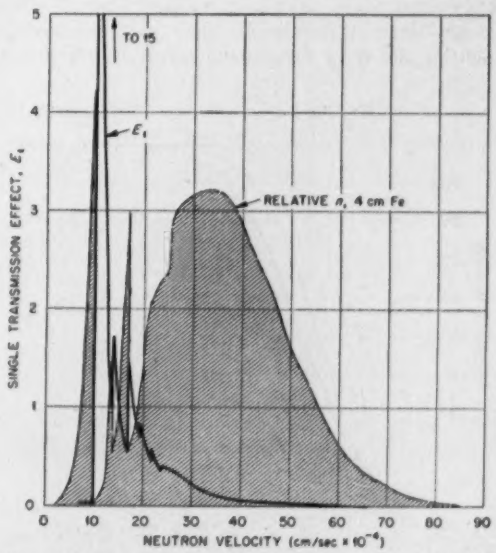

FIG. 3. Single transmission effect vs velocity for 4-cm block using measured value of the depolarization coefficient and Eq. (1). The neutron density n emerging from 4 cm of iron is shown by the dashed curve.

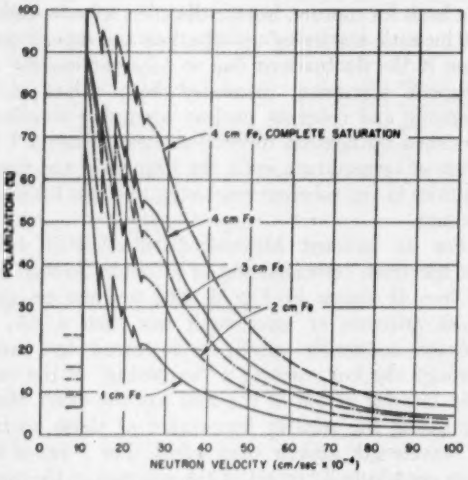

FIG. 4. Polarization *vs* velocity for 1, 2, 3, and 4 cm of iron.

the magnet. In order to keep the spectrum incident upon the detector always the same, a total of 4 cm of iron was always kept in the path of the beam. The single transmission effect is the ratio of the counting rate with d cm of magnetized iron, and $(4d)$ cm of unmagnetized iron in the beam, to that with 4 cm of unmagnetized iron in the beam. [The magnetic field at the surface of the magnetized iron was always kept at the value used in the nuclear polarization experiment[22] (11 000 oersteds).]

The measured values of the single transmission effect for the entire spectrum as a function of block thickness were compared to calculated values. The calculated average single transmission effect was obtained by weighting the value for a single energy by the density

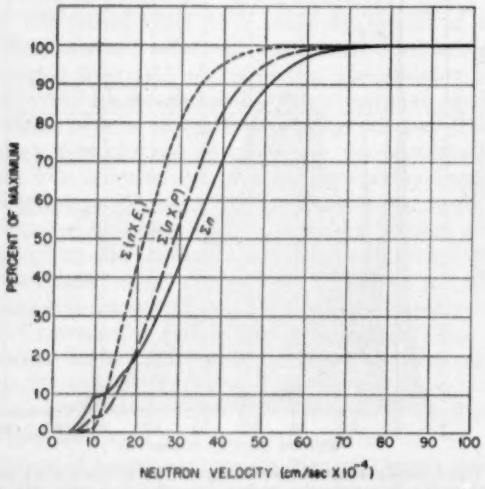

FIG. 5. Cumulative plots of neutron density, neutron density×E_1, and neutron density×polarization for 4 cm of Fe.

of neutrons in the beam emerging from 4 cm of iron. The integrations were performed numerically. The value for the depolarization coefficient determined previously was used. It so happened that a good fit to the data was obtained using the boundary condition that the slope of the $3d$ wave function is zero at the atomic radius with the value 12.3 barns for the coherent scattering cross section of iron. This choice is obviously not unique. The value 10.0 barns and the alternate boundary condition would do as well. For wavelengths greater than about 1.5A the values of the form factor given in reference 20 were used. Reference 20 gives no values of $F(l)$ for $l^2>14$, which are needed in addition to those for $l^2<14$ in calculating the polarization cross section for wavelengths less than 1.5A. These were estimated from a brief approximate analytical expression[26] giving F for a free iron atom. The error involved in using these rough values of F for $l^2>14$ is small, since a large fraction of the polarization cross section at the short wavelengths is contributed by planes for which $l^2<14$.

Those neutrons of wavelength greater than $\lambda=2d_{110}$ or 4.04A are not scattered coherently from iron. Those neutrons of this energy region which are scattered inelastically from magnetized iron may be polarized.[27] We have used measured values of the single transmission effect in this region[28] to calculate the contribution to the polarization of these low-energy neutrons. For 4 cm of iron, 8 percent of the neutrons emerging are below 4.04A. These neutrons have a polarization of about 16 percent.

The single transmission effect for 4 cm of iron is shown as a function of velocity in Fig. 3. These values were calculated using (1). The large values of the single transmission effect at velocities slightly greater than that of the coherent scattering cutoff of iron emphasize the possible importance of these neutrons in a weighted average.

The polarization as a function of velocity calculated from (1), (2), and (4) is shown in Fig. 4 for 1-, 2-, 3-, and 4-cm thicknesses of iron. The polarization *vs* velocity for 4 cm with complete saturation is shown for comparison. Cumulative plots against velocity of neutron density×single transmission effect and neutron density×polarization for 4 cm are shown in Fig. 5 along with the cumulative neutrun density as a function of velocity. From these plots the relative contributions of selected portions of the spectrum can be taken. The curve $\sum(n\times E_1)$ rises more quickly than the curve $\sum(n\times P)$, because to a first approximation, the single transmission effect is proportional to p^2, while the polarization is proportional to p.

The curves described above were used in calculating the single transmission effect for a $1/v$ detector weighted over the emerging spectrum through 1, 2, 3, and 4 cm

[26] Hoffman, Livingston, and Bethe, Phys. Rev. **55**, 924 (1939).
[27] O. Halpern, Phys. Rev. **72**, 261 (1947).
[28] Hughes, Burgy, and Woolf, Phys. Rev. **80**, 481 (1950).

of iron for all velocities from zero up to the Cd cutoff. The calculated values are compared with the measured values of Table III. The measured and calculated values agree to about 0.5 percent.

The measured values of single transmission effect having been acceptably accounted for, the curves were then used to calculate a polarization weighted with neutron density over the spectrum. The average polarization for energies greater than the iron cutoff and less than the Cd cutoff (0.5 ev) was 37 percent. The average polarization from zero energy up to the Cd cutoff is 35 percent.

The epi-cadmium neutrons are polarized only to a small degree. Their effect upon the average polarization is small because only about 10 percent of the activity of the sample used in the nuclear polarization experiment was due to epi-cadmium neutrons. We estimate that the 35 percent average polarization of all those neutrons below 0.5 ev becomes 32 percent when the averaging is done over the entire spectrum.

The polarization cross section corresponding to the average single transmission effect 0.221 for all energies less than 0.5 ev for the magnetic state of saturation of

TABLE III. Single transmission of continuous spectrum with $1/v$ detector.

Thickness cm	(j/j_0), measured	(j/j_0), calculated
1	1.022±0.001	1.016
2	1.063±0.001	1.059
3	1.136±0.001	1.130
4	1.221±0.001	1.231

our experiment in which $q=0.2$, is 2.2 barns. This value of p corresponds to a polarization of 45 percent under the same conditions of magnetization. The average polarization for all neutrons in this energy range is actually 35 percent. These numbers illustrate the degree of accuracy to be expected in deducing the average polarization from the average single transmission effect without taking into account the details of the spectrum.

E. MEASUREMENT OF POLARIZATION USING MAGNETITE

The polarization of a beam of neutrons of a single energy produced by a given polarizer can easily be measured if the polarization produced by a second polarizer from the same incident unpolarized beam is known. A schematic arrangement of the apparatus is shown in Fig. 6. The incident beam of unpolarized neutrons is passed through the polarizer A and analyzer B. The counting rate C is measured with A and B in parallel so that they both discriminate against neutrons of the same orientation. The counting rate C_2 obtained when the polarization of the beam between A and B is brought to zero by passing it through a shim of iron placed in zero field, is also measured. It can be

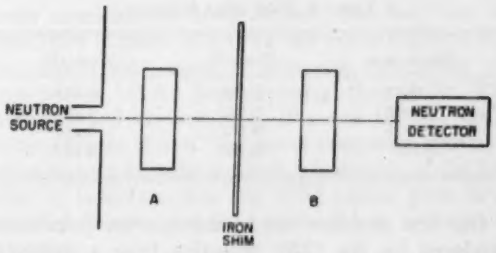

FIG. 6. Schematic arrangement for measuring polarization. A and B are polarizing regions. The shim reduces the polarization to zero.

shown that

$$G=C_1/C_2=1+P_1P_2, \tag{7}$$

where P_i is the polarization produced by polarizer i alone from the incident unpolarized beam of neutrons. We call G the "shim effect." This method has the advantage that $(G-1)$ is linear with the unknown polarization. It is equally sensitive for small and large polarizations. Measurements of small values of the polarization by the single transmission method are very difficult to make with accuracy as can be seen from Fig. 1. A polarization of 10 percent gives a single transmission effect of only 0.005.

We have attempted to measure unknown polarizations by using the polarization properties of the (220) reflection from a single crystal of magnetite. The polarization produced by a given set of planes of Miller indices (hkl), can be calculated[12] from the known chemical and magnetic structure of magnetite, the known values of the relevant nuclear and magnetic scattering amplitudes, and the direction of the applied magnetic field with respect to the scattering vector. For the applied field perpendicular to the scattering plane, the polarization of the (220) reflection is calculated to be close to 100 percent. The polarization of the second order (440) is calculated to be about minus 10 percent, the negative sign denoting polarization in the direction opposite to that of the first order (220). For the (220) reflection of the 56 atoms per unit cell, only the eight Fe^{+++} ions in the tetrahedral positions contribute to the coherent scattering. All 56 atoms contribute to the (440) reflection. Consequently, a crystal set to reflect a 100 percent polarized beam in the first order from the (220) planes, will reflect also in the same direction an appreciable intensity of second-order neutrons polarized slightly in the opposite direction.

TABLE IV. Shim effect for magnetite crystals.

Crystal in position A	Crystal in position B	Shim effect
III	I	1.430
III	II	1.437
I	II	1.547
I	III	1.385
I	IV	1.353

TABLE V. Shim effect *vs* energy.

Energy in ev	$(G_{obs}-1)$	$(G_{corr}-1)$
0.0372	0.138	0.369
0.0600	0.192	0.277
0.0858	0.177	0.215
0.149	0.103	0.132

Our first problem was to measure the polarization produced by the (220) reflection from a magnetite crystal. We procured several large natural crystals, and cut out several slabs $\frac{1}{8}$ in. thick and of area about 2 square in. in the (220) plane. The polarizations produced by the slabs were compared by measuring the shim effect with one of the slabs in position *A*, and with each of the other slabs in turn in position *B*. The results as shown in Table IV.

As can be seen from the data, the polarizations produced by the crystals were in general not the same. Crystal slabs I and II which were cut from the same crystal gave closely the same shim effect. If $P_1=P_2$, then $G-1=P^2$. The measured shim effect of 1.55 using crystals I and II together was corrected for higher-order content to give a value for first order only of 1.81. The polarization produced by each of these crystals alone is, therefore, about 90 percent. The polarization produced by crystal slabs III or IV alone would have been less than 90 percent, as indicated by their lower values for the shim effect using I in position *A*. Ordered impurities in the spinel structure of our large crystals are perhaps responsible for the values of the polarization being less than that expected on the basis of measured scattering amplitudes and the crystal structure.

We have used the value of 90 percent for the polarization produced by crystal I alone, to measure the values for the polarization in the beam coming through 4 cm of magnetized iron as a function of energy using expression (7). The shim effect was measured and corrected for the presence of second order by use of the boron transmission measurements. Values for the polarization cross section p were then derived for these energies from the polarization of the beam from the iron, using the previously measured depolarization coefficient in expressions (1), (2), and (4). These values of p are shown in Fig. 2.

The great disadvantage of the use of magnetite as analyzer is the large second-order correction required

TABLE VI. Polarization *vs* energy.

Energy in ev	Polarization as measured by transmission, percent	Polarization as measured with magnetite, percent
0.0372	49	41
0.0600	35	31
0.0858	26	24
0.149	13	15

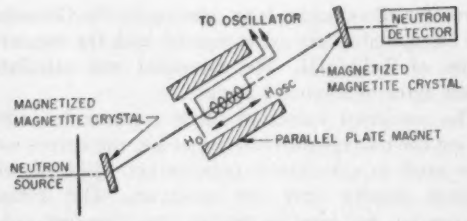

FIG. 7. Arrangement for measuring polarization using two magnetite crystals and magnetic resonance method.

in the measured value of $(G-1)$. The observed and corrected values for the quantity $(G-1)$ are given in Table V.

Values of the polarization at four energies deduced from the single transmission and magnetite methods are compared in Table VI. The agreement is quite good for all but the lowest energy 0.0372 ev, where the correction for higher order in the magnetite method is unacceptably large. For this energy, we take the transmission method to give more reliable results. The method described for measuring unknown polarization using a magnetic crystal of known properties as analyzer would be more practical if a crystal giving small higher-order content along with high neutron polarization were available.

It was pointed out above that the shim effect $G=1+P_1P_2$ is a more sensitive measure of polarization than the single transmission effect. An even more sensitive measure is the double transmission effect,[10,16] the ratio of the intensity with polarizer and analyzer aiding to that with polarizer and analyzer opposing, which is given by

$$K=(1+P_1P_2)/(1-P_1P_2). \tag{8}$$

We have used the arrangement of Fig. 7 to obtain values of K. The reversal of P_1 was achieved in the region between crystals I and II by the magnetic resonance method,[29] in which the steady field H_0 was perpendicular to the direction of the beam, and an oscillating field H_{osc} was along the beam direction. Polarizer, I, and analyzer, II, are again our magnetized magnetite crystals, in which $P_1=P_2=0.9$. Intensity measurements were taken under the four sets of conditions listed in Table VII. The shim referred to in Table VII is a depolarizing sheet of iron in zero field. Turning the oscillator on, changes the intensity from 100 to 30.2. BF_3 absorption measurements indicate that a large

TABLE VII. Intensity measurements using oscillator.

Conditions	Relative intensity
1. Oscillator off	100
2. Oscillator on	30.2
3. Oscillator on, with shim	64.8
4. Oscillator off, with shim	64.8

[29] L. W. Alvarez and F. Bloch, Phys. Rev. **57**, 111 (1939).

fraction of the residual 30.2 is due to the presence of second-order neutrons. The oscillating field affects the second-order neutrons very little for two reasons: K will be close to one because their polarization is low (-10 percent); when the strength of the oscillating field is such that the probability of flipping first-order neutrons is a maximum, the probability of flipping second-order neutrons is considerably less than maximum. The symmetry of the intensities in cases (3) and (1) with respect to (2) indicates strongly that all first-order neutrons are being flipped. The fact that the intensities in cases (3) and (4) are equal indicates that the depolarizing shim is removing completely the polarization of the beam passing through it. This method could be used, also, if I and II were transmitting blocks of iron. The use of magnetite crystals is advantageous because the high polarization of the first order of reflection from the (220) planes gives large intensity changes, and the polarization is independent of energy.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

The Alpha- and Gamma-Ray Spectra of Pu^{238}†

FRANK ASARO AND I. PERLMAN
Radiation Laboratory and Department of Chemistry, University of California, Berkeley, California
(Received January 12, 1954)

The alpha and gamma spectra of Pu^{238} have been studied with an alpha-particle spectrograph and gamma-ray scintillation and proportional counters. Alpha groups of 5.495 (72 percent), 5.452 (28 percent), and 5.352 Mev (0.09 percent) and electromagnetic radiations of 17 (13 percent), 43.8±0.5 (0.038 percent), 99±2 (0.008 percent), and 150±2 kev (0.001 percent) were observed. Spins and parities are assigned to the energy levels, and the results are evaluated with respect to the developing theory and systematics of complex alpha spectra and excited states of even-even nuclei.

I. INTRODUCTION

MARKED regularities have been observed in the alpha spectra of different *even-even* nuclei indicating strongly that corresponding spectroscopic states are involved in the decay processes. Among the heaviest nuclei the low-lying excited states are spaced in such a fashion as to suggest that they constitute a rotational band[1] interpreted according to the Bohr and Mottelson[2] theory which unifies independent particle and collective aspects of nuclear states.

In another publication[3] the details of the alpha spectrum and gamma spectrum of Cm^{242} were reported and discussed in terms of the excited states of Pu^{238}. The present communication deals with Pu^{238} and the energy levels of U^{234}. The close similarity between Cm^{242} and Pu^{238} spectra can be seen from the decay schemes shown in Fig. 1, and it will be seen that the gamma-ray conversion coefficient data also show that comparable spectroscopic states are involved.

II. METHODS

The samples of Pu^{238} used in the present measurements[4] were made by the prolonged neutron irradiation of Am^{241} through the following series of reactions:

$$Am^{241}(n,\gamma)\,Am^{242m} \xrightarrow[16\text{ hr}]{\beta} Cm^{242},$$

$$Cm^{242} \xrightarrow[162\text{ day}]{\alpha} Pu^{238}.$$

The primary objective of the irradiations was to make Pu^{242} through the electron capture branching of 16-hour Am^{242m} so sizable amounts of this isotope were present. Also in a long irradiation the Pu^{238} captures neutrons to give successively Pu^{239}, Pu^{240}, and Pu^{241}, all of which were also present in small quantities. Mass spectrometric analyses[5] were made on two of the preparations

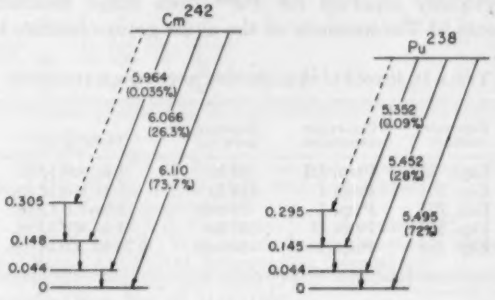

FIG. 1. Comparison of alpha spectra of Cm^{242} and Pu^{238}.

† This work was performed under the auspices of the U. S. Atomic Energy Commission.

[1] F. Asaro and I. Perlman, Phys. Rev. **91**, 763 (1953).

[2] A. Bohr and B. R. Mottelson, Phys. Rev. **89**, 316 (1953); **90**, 717 (1953); Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **27**, No. 16 (1953).

[3] Asaro, Thompson, and Perlman, Phys. Rev. **92**, 694 (1953).

[4] We are indebted to Dr. S. G. Thompson of this laboratory and to the personnel of the Atomic Energy of Canada, Ltd., Chalk River, Ontario, Canada, for making some of these preparations available to us.

[5] The analyses were made by Mr. F. L. Reynolds of this laboratory.

TABLE I. Mass composition of plutonium preparations.

	238	239	240	241	242
Prep. I	62.9%	11.7%	1.6%	0.07%	24%
Prep. II	1.5%	80.0%	4.2%	0.5 %	13.7%

and the results are shown in Table I. The large amount of Pu^{239} in Preparation II was the result of Pu^{239} present in the Am^{241} before irradiation. Since Pu^{238} has a half-life which is short relative to the alpha-decay half-lives of the other isotopes, the alpha activity in both preparations was predominantly that of Pu^{238}. In Preparation I the Pu^{238} comprised 99.9 percent of the alpha activity and in Preparation II it was 81 percent. However, the differences in composition were used to assign the rarer alpha groups.

A third sample (Preparation III) of Pu^{238} was not mass analyzed but should have been free of other plutonium isotopes since it came from some previously purified curium (Cm^{242}) which had decayed through approximately one half-life.

All samples were mounted by vacuum sublimation of plutonium chloride onto a platinum plate which was masked to present a band 1 in.$\times\frac{1}{8}$ in. This technique as well as the spectrograph description and operation have appeared in other reports.[6,5] As before, the alpha particles were caught on a photographic plate and the track count plotted according to position on the plate in order to reproduce the spectrum. The record of alpha-particle spectrograph exposures is given in Table II.

Gamma-ray spectra were measured for the most part with a sodium iodide scintillation counter coupled with either a single channel or multichannel pulse-height analyzer. In some experiments a xenon-filled proportional counter was used to produce the pulse for the analyzer.

III. RESULTS

A. Principal Alpha Groups of Pu^{238}

A sample of Cm^{242} which had decayed six months was run on the spectrometer (Prep. III, Exp. 73) and showed two new alpha groups at about the energy previously reported for Pu^{238} from range measurements.[7,8] The intensity of the alpha groups relative to Cm^{242} groups was 0.6 percent in agreement with expectations for the particular growth and decay period. The main alpha group of Cm^{242} was used as an energy standard (6.110 Mev),[8] and the value for the main group of Pu^{238} was found to be 5.492 Mev. The second group was of 42 kev lower energy and therefore 5.450 Mev.

TABLE II. Record of alpha-particle spectrograph exposures.

Exposure number	Plutonium preparation	Exposure duration	Defining slit
Exp. 73	Prep. III	40 hr	1 in.×0.12 in.
Exp. 204	Prep. I	114 hr	1 in.×0.12 in.
Exp. 205	Prep. I	75 min	1 in.×0.12 in.
Exp. 243	Prep. II	37 hr	1 in.×0.12 in.
Exp. 283	Prep. I	54 min	0.75 in.×0.018 in.

The curium of Prep. III could have contained Am^{241} which has a prominent alpha group at 5.476 Mev which could distort the energy determination of the Pu^{238} groups. However, it was found that the sample contained less than 2 percent by weight of Am^{241}, and its alpha group would therefore be at least 300-fold less intense than the Pu^{238} group.

The particle energy of Pu^{238} was also measured in Exp. 243, Prep. II, using the 5.150-Mev group[9] of Pu^{239} as a standard. This gave 5.497 Mev for the main group of Pu^{238} and a 42.5-kev separation for the second group. In Exp. 283 the separation between the groups was again 42.5 kev. We shall take as the best value for the main group 5.495 Mev and 5.453 Mev for the second group.

These groups presumably lead, respectively, to the ground state and first-excited state of U^{234}. Applying the recoil correction to the energy spearation we find that the first-excited state of U^{234} is 43 kev above the ground state. If we use the terminology described in previous publications the alpha group populating the ground state will be designated α_0 and that populating the 43-kev state α_{43}. The intensity of α_{43} relative to total Pu^{238} alpha particles was found to be 24 percent in Exp. 73 and 28 percent in Exp. 283.

B. Low-Intensity Alpha Groups of Pu^{238}

In Exp. 243 (Prep. II) and Exp. 204 (Prep. I) a third group of Pu^{238} was observed. The separation in particle energy between this group and α_{43} of Pu^{238} was found to be 101.5 kev in Exp. 243 and 98.6 kev in Exp. 204. The average value of 100 kev is accepted which means that the energy level of U^{234} is 102 kev above the 43-kev level and 145 kev above the ground state.

Although α_{145} is in very low intensity (0.1 percent of total Pu^{238}), it was proved to belong to Pu^{238} because of its constant relationship with the Pu^{238} alpha particles in the two preparations which differed widely in isotopic composition. This information is shown in Table III.

The exact correspondence of energy levels and alpha transitions to these levels between different even-even nuclei in this region is worthy of note. As shown in Fig. 1, the alpha spectra of Cm^{242} and Pu^{238} point to energy levels of the product nuclei spaced almost identically. Each has a transition to the ground state and to the first-excited state whose intensities can be shown to be in substantial agreement with alpha-decay theory, and each has a transition to the second-excited

[6] Asaro, Reynolds, and Perlman, Phys. Rev. **87**, 277 (1952).
[7] W. Thiel and A. H. Jaffey, reported in Argonne National Laboratory Report ANL-4370, November, 1949 (unpublished).
[8] Chamberlain, Gofman, Segrè, and Wahl, Phys. Rev. **71**, 529 (1947).
[9] F. Asaro and I. Perlman, Phys. Rev. **88**, 828 (1952).

state which is highly hindered (low intensity far beyond expectations according to the energy difference).

C. Gamma Rays of Pu^{238}

The results to be presented include the energies and intensities of the photons accompanying the alpha decay of Pu^{238}. The best picture which can be constructed from these data is that all of the transitions are of the $E2$ type and that they cascade as indicated in Fig. 1. It will be noted that the gamma ray of 150 kev is shown to come from a level which is not populated by an observable alpha group. A similar assignment was made for Cm^{242} and the reasons will be presented below.

A typical gamma-ray spectrum taken with a sodium iodide crystal spectrometer is shown in Fig. 2. The relative intensities of the photons are not directly discernible because absorbers were used to minimize the intense L x-rays (18 kev) which would otherwise obscure the 42-kev peak.

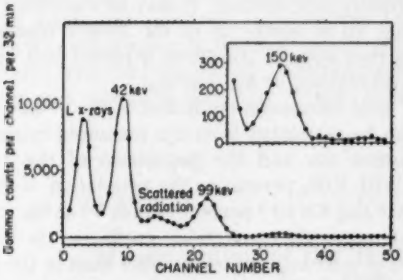

FIG. 2. Pu^{238} gamma spectrum.

In all, ten measurements of the gamma-ray spectrum were made; nine of these were done on Prep. I and for the other a different sample of Pu^{238} was used. The experiments aimed at energy determination are summarized in Table IV. The intensities will be discussed presently.

In experiments 1, 2, and 3 the voltage on the photomultiplier tube was not stable and a single channel analyzer was employed. These energy values are not considered as reliable as the others. Experiments 4, 5, and 6 were run with multichannel analyzers and under more stable conditions.

D. The 43.8-kev Gamma Ray

A xenon-filled proportional counter was used in experiments 7 and 8 and should give the most reliable values for the ~40-kev gamma ray. For these measurements the counter was standardized with the 44.3-kev gamma ray[9] of Cm^{242}. This value (44.3 kev) obtained from the energy difference of alpha groups is in good agreement with the energy obtained by measuring the photon in a proportional counter (44.6 kev),[10] from the conversion electron spectrum (44.1 kev),[11] and from the conversion electron spectrum in the decay of Np^{238} (44.1 kev).[12] The gamma-ray energy for Pu^{238} so obtained was 43.8±0.5 kev.

TABLE III. Ratios of intensities of $Pu^{238}\alpha_{145}$ and other alpha groups.

Intensity relative to:	Exp. 204	Exp. 243
Pu^{238} alpha activity	0.09$_2$%	0.09$_9$%
Pu^{239} alpha activity	130%	0.5%
Pu^{240} alpha activity	265%	2.6%
Pu^{241} beta activity	13%	0.05%
Pu^{242} alpha activity	1300%	61%

The intensity of this gamma ray relative to total alpha particles of Pu^{238} was determined by Reed[13] using a xenon-filled Geiger tube and absorbers. He found the intensity to be 4.2×10^{-3}. In the present study (experiment 2) the intensity was measured with the scintillation spectrometer using the 60-kev gamma ray of Am^{241} as a standard. The intensity of the Am^{241} gamma ray is 0.40 per disintegration as measured by Beling, Newton, and Rose.[14] The intensity of the gamma ray so obtained was 3.8×10^{-4} which is considerably lower than Reed's measurement and should be more accurate because of the inherent better precision in the method used.

There can be no question that this gamma ray results from the de-excitation of the state reached by Pu^{238} α_{43} (see Fig. 1.) Since α_{43} populates this state to the extent of 28 percent, the low photon intensity would imply a large conversion coefficient. O'Kelley[15] has observed in good intensity a series of conversion lines attributable to a transition of ~45 kev and Dunlavey and Seaborg[16] found by the photographic plate technique that 23 percent of Pu^{238} alpha particles are in coincidence with electrons corresponding to an ~40-kev gamma ray. Within experimental uncertainties this number (23 percent) is the same as the intensity of α_{43} (28 percent)

TABLE IV. Energies of Pu^{238} electromagnetic radiations.

Experiment No.										Best
1	2	3	4	5	6	7	8	9	10	value
157[a]	...	145	146	148	150±2	...	...	...	151	150
107	102	99	101	96	99±2	...	...	...	97	99
45	42	...	...	...	42±2	43±1	43.8±0.5	...	41	43.8
17	18	...	...	...	18	...	...	17±0.5	...	17

[a] Energies in kev.

[9] B. Rose and J. O. Newton, reported at Birmingham Conference on Nuclear Physics, July, 1953 (unpublished).
[10] T. O. Passell (unpublished).
[11] H. Slätis and J. O. Rasmussen, Jr. (to be published).
[12] G. W. Reed, Jr., U. S. Atomic Energy Commission Declassified Report AECD-3185, January, 1951 (unpublished).
[13] Beling, Newton, and Rose, Phys. Rev. 86, 797 (1952); erratum: Phys. Rev. 87, 1144 (1952).
[14] G. D. O'Kelley, Ph.D. thesis, University of California Radiation Laboratory Report UCRL-1243, March 15, 1951 (unpublished).
[15] D. C. Dunlavey and G. T. Seaborg, Phys. Rev. 87, 165 (1952).

indicating a large conversion coefficient. With 28 percent as the population of the 44-kev state and 0.038 percent for the photon intensity, the total conversion coefficient is 740. In order to determine the L-conversion coefficient one must know the relative amounts of conversion in the L and higher-order shells. For a similar transition in the decay of Cm^{242} it was found that 80 percent of the conversion electrons were from the L shell and 20 percent from M and higher shells. Applying these values to the Pu^{238} case, the L-shell conversion coefficient becomes 600. This can be compared with theoretical calculations of Gellman, Griffith, and Stanley[17] from which one would expect values of 1, 60, and 700 for $E1$, $M1$, and $E2$ transitions, respectively. The transition is obviously of $E2$ type on this basis and this is added evidence that the first-excited state of an even-even nucleus has spin 2 and even parity.

E. L X-Rays

Photons of 17–18 kev were the most prominent electromagnetic radiation in the decay of Pu^{238}. The best energy value was 17 ± 0.5 kev determined in experiment 9 (Table IV) with the scintillation spectrometer and a motor-driven single channel analyzer. These are undoubtedly uranium L x-rays which arise almost entirely from the internal conversion of the 43.8-kev gamma ray.

The intensity of this L x-ray complex was obtained by comparison with the L x-rays of Am^{241} decay, the intensity of which was determined by Beling, Newton, and Rose.[14] The value so obtained was 13 percent (Table V) and this may be compared with the expectations from the internal conversion of the 43.8-kev gamma ray.

Since the 43.8-kev gamma ray is about 80 percent converted in the L shell (see previous discussion) there should be 0.22 L vacancies per alpha disintegration. The calculations of Gellman, Griffith, and Stanley[17] show that an $E2$ transition such as this is converted almost entirely in the L_{II} and L_{III} shells and Kinsey[18] has estimated that the fluorescent yield for these L vacancies is ~0.5. The expected number of L x-rays is then 11 percent (0.5×22) which is in satisfactory agreement with the observed intensity, 13 percent.

TABLE V. Intensities of Pu^{238} electromagnetic radiations.

Energy (kev)	Absolute intensity per alpha disintegration	Intensity relative to 43.8-kev radiation	Intensity relative to 99-kev radiation	Experiment number	Best value for absolute intensity per alpha disintegration
17	13%			9	13%
43.8	0.038%			2	0.038%
99		22%		2	
					0.0080%
		20%		10	
150			11.8%	1	
			12.8%	4	
					0.0010%
			12.6%	5	
			13%	10	

[17] Gellman, Griffith and Stanley, Phys. Rev. 85, 944 (1952).
[18] B. B. Kinsey, Can. J. Research. A26, 404 (1948).

F. The 99-kev Gamma Ray

As indicated in Fig. 1, this gamma ray has been placed between the states populated by α_{145} and α_{44}. It will be noted that a gamma ray of 150 kev was also observed and it might be assumed that this is the crossover transition from the state reached by α_{145} to the ground state and that the 99-kev photon is the K x-ray resulting from internal conversion of the 150-kev gamma ray. From the relative intensities of the 99-kev and 150-kev photons the K-conversion coefficient would have to be 8. Furthermore, all transitions to the ground state of an even-even nucleus from levels populated by alpha decay must be electric. From the data on K-shell conversion coefficients by Rose and co-workers[19] and by Spinrad and Keller,[20] an electric transition of 150 kev with conversion coefficient 8 would have to have a multipolarity of at least 5 and would therefore have an extremely long lifetime. It may be concluded therefore that all or nearly all of the 99-kev photons are gamma rays and the transition is placed best between the levels reached by α_{145} and α_{44}.

The total conversion coefficient of the 99-kev transition can be calculated from the measured intensity of the gamma ray and the population of the 145-kev state. With 0.09_6 percent as the population of the 145-kev state and 8×10^{-3} percent (Table V) as the gamma-ray intensity, the conversion coefficient is 11. The amount of internal conversion other than in the L shell is probably small so that this figure may be taken as approximately the L-conversion coefficient. From the tables of Gellman, Griffith, and Stanley[17] the theoretical values for $E1$, $E2$, and $M1$ transitions are respectively 0.1, 8, and 4. The $E1$ transition is ruled out, and, although $E2$ gives a better fit than $M1$, a choice probably should not be made on this ground alone. Since the 44-kev state has even parity (2+), the 145-kev state must therefore also have even parity. Only even spin numbers are allowed since the level is reached by alpha decay of an even-even nucleus. The possible states are therefore 0+, 2+, and 4+, and of these the selection of 4+ is favored for reasons to be given below.

G. The 150-kev Gamma Ray

The low intensity photon of 150 kev (see Fig. 1 and Tables IV and V) has already been mentioned in the preceding section where it was shown that it cannot account for much of the de-excitation of the 145-kev state. Gamma-gamma coincidence measurements showed this gamma ray to be in coincidence with the

[19] Rose, Goertzel, Spinrad, Harr, and Strong, Phys. Rev. 83, 79 (1951); Rose, Goertzel, and Perry, Oak Ridge National Laboratory Report ORNL-1023, June, 1951 (unpublished).
[20] B. I. Spinrad and L. B. Keller, Phys. Rev. 84, 1056 (1951).

99-kev photon. Furthermore, the coincidence rate was in agreement with the assignment of the 99-kev photon to a gamma ray with conversion coefficient ~10.

As will be borne out in the discussion, it is reasonable to assign the 150-kev transition to another of the series of cascading $E2$ transitions (see Fig. 1). On this basis, its conversion coefficient would be 2 and the population of the 295-kev level would be 3×10^{-3} percent. This intensity was below the limits of our detection of an alpha group which for this case was $\sim10^{-2}$ percent, so the failure to observe the corresponding alpha group is not inconsistent with the decay scheme.

IV. DISCUSSION

The energy levels of U^{234} observed from the alpha decay of Pu^{238} conform in remarkable fashion with the picture for even-even nuclei well removed from a closed shell developed by Bohr and Mottelson.[2] In this model collective aspects of nuclear structure and individual particle aspects are coupled and in a region well removed from a closed shell there should be a series of energy levels corresponding to a rotational band in which only even spin states (0+, 2+, 4+, ···) appear. On this basis electric quadrupole transitions should predominate. It was seen that the 44-kev transition definitely falls into this category and the 99-kev transition probably does so. A unique assignment could not be made for the 150-kev gamma ray because the conversion coefficient could not be measured, but the failure to observe an alpha particle leading to the 295-kev state means that the conversion coefficient is not an order of magnitude greater than unity and is therefore in the proper range for an $E2$ transition.

Another aspect of the Bohr and Mottelson model concerns the energy level spacings of the states 0, 2, 4, ···. The spectrum is that of a rigid rotator although it should be pointed out that in the treatment the moment of inertia is not that of a rigid body but is an effective moment associated with the angular momentum of the surface waves. The rotational energy levels go as $I(I+1)$, where I is the spin number so the energy levels go as 1, 3.3, 7.0 for the first three excited states of spin 2, 4, 6. In this case, the ratio between the second- and first-excited states (145/44) is 3.3 and between the third and first (295/44) is 6.7. This striking agreement with the theory is the principal reason for assigning the second- and third-excited states 4+ and 6+, respectively, in the absence of independent means of making assignments.

We wish to acknowledge the assistance of Mrs. J. A. Simmons in counting the alpha tracks.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Measurement of the Tritium Concentration in Natural Waters by a Diffusion Cloud Chamber*

E. L. Fireman and D. Schwarzer
Brookhaven National Laboratory, Upton, New York

(Received January 8, 1954)

The radioactivity of tritium is easily recognized and visually counted in a hydrogen-filled diffusion chamber. A 6-in. diameter diffusion chamber, filled with 100 psi of hydrogen obtained by completely converting 14 cc of water in a Mg furnace, is used to measure the tritium concentration in surface water of the Long Island Sound, in New Jersey well water, and in a number of rain and atmospheric moisture samples. These samples had tritium concentration (T/H ratios) that ranged from 10^{-15} to 3×10^{-18}. Some samples were measured directly, others required some electrolysis to give measurable tritium concentrations. No tritium could be detected in glacial water.

INTRODUCTION

THE existence of natural tritium in Norwegian lake water has been demonstrated.[1] This was accomplished by enriching the natural tritium content by electrolysis between 1 and 10 million fold. A series of measurements on natural tritium in rain and ground water is being carried out at the University of Chicago[2] with a low-level Geiger counter and an electrolytic plant. We have applied a diffusion chamber technique to the same problem. This technique is more sensitive than a low-level Geiger counter[3] that we have been using here. Therefore, considerably less enrichment is required to measure the tritium in natural waters. In fact, in two rain samples and one atmospheric moisture sample the tritium could be measured directly without any enrichment.

DESCRIPTION OF APPARATUS

Figure 1 is a schematic diagram of the diffusion chamber and its associated hydrogen-converting ap-

* Research performed under the auspices of the U. S. Atomic Energy Commission.

[1] Grosse, Johnston, Wolfgang, and Libby, Science **113**, 1 (1951).

[2] W. F. Libby, Phys. Rev. **93**, 1337 (1954); Proc. Natl. Acad. Sci. **39**, 245 (1953).

[3] E. L. Fireman, Phys. Rev. **91**, 922 (1953).

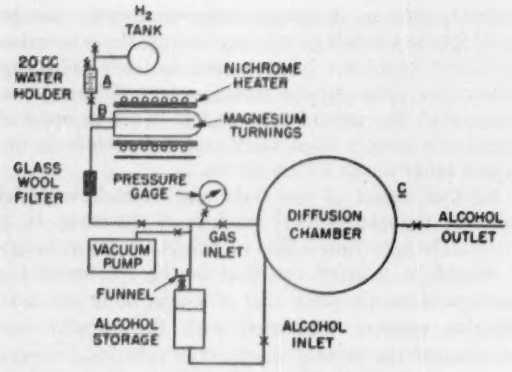

FIG. 1. Schematic block diagram.

paratus. This equipment takes up little space, is easy to build, requires no shielding, requires no electronic equipment and less than 20 amperes from a 110-volt wall plug.

The chamber is filled by putting the water sample into a 20-cc medical syringe, labeled *A* in Fig. 1. Pressure from a standard H_2 tank forces the water into Mg furnace, *B*, which had previously been evacuated and heated to 650°C. The hydrogen evolved passes through a glass wool filter into the diffusion chamber. A cold trap originally in the line was removed as unnecessary because no water was trapped. Stoichoimetry gives a good check on the completeness of the conversion. The complete conversion of 14 cc of water fills the unit to 100-psi pressure. The Mg turnings in the furnace are replaced after four chamber fillings. Figure 2 is a drawing of the Mg furnace.

After a measurement that shows little tritium activity the furnace and cloud chamber are merely evacuated before the next filling. However, if more than 10 tritium counts/minute appear in the chamber it is necessary to convert 5 cc of nonactive water in the furnace and flush the chamber with tank H_2 in order to eliminate traces of tritium.

The conversion of 14 cc of water requires about 15 minutes time. Visual counting can be started 30 minutes later when the chamber reaches temperature equilibrium.

Figure 3 is a cross section of the diffusion chamber. It consists of a 6-in. diameter Pyrex cylinder, 3 in. high and $\frac{1}{4}$ in. thick, between a stainless-steel chamber bottom and steel ring acting as a central temperature control. A 1-in. Bakelite ring insulates

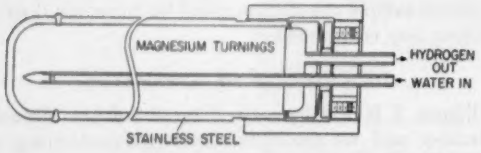

FIG. 2. Magnesium furnace.

the alcohol tray from the central ring. Above this combination of cylinder and rings is the viewing window consisting of a $\frac{1}{2}$-in. sheet of Allite and a 1-in. sheet of Lucite. The whole thing is clamped together by twelve $\frac{1}{4}$-in. diameter stainless steel studs.

The chamber bottom is cooled to −65°C by circulating triclorethylene at dry ice temperature through copper tubing that is soldered to the chamber bottom. Black dye in alcohol on the chamber floor serves as an excellent viewing background. The height of the sensitive layer is almost the same as that of the Pyrex cylinder. The central temperature ring is cooled to −25°C by connecting it to the bottom with 24 copper rods of $\frac{3}{16}$-in. diameter. A Lucite cylinder surrounds

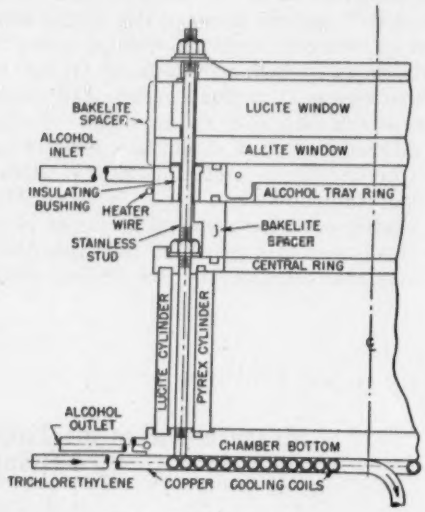

FIG. 3. Diffusion chamber.

the copper rods and Pyrex cylinder and insulates them from the outside. A heater wire about the alcohol tray heats it to +10°C. The above temperatures are not very critical; a few degrees variation does not affect the chamber operation.

ELECTROLYSIS

The tritium in some water samples is concentrated by electrolysis. The procedure described by Urey[4] is followed. Two nickel cells are being used. One concentrates 250 cc of water with 5 g of NaOH to 20 cc. The other concentrates 1500 cc with 30 g of NaOH to 200 cc. A conventional 75-ampere battery charger supplies sufficient current and voltage. The cells are cooled to about 55°F with tap water; no provision is made to catch escaping water vapor. The calibration of the tritium enrichment is done with calibrated tritiated water and with rain whose tritium concentration can be measured directly. The fractionation factor

[4] H. C. Urey and G. K. Teal, Revs. Modern Phys. 7, 34 (1935).

under these conditions is (6.9±1.0), so that the electrolysis of 1500 cc of water to 20 cc increases the tritium concentration by a factor of 40. In the measurements carried out so far, no sample was enriched by more than a factor of 40.

VISUAL COUNTING METHOD

The tritium tracks are short electron tracks with ranges of $\frac{1}{8}$ in. or less in the chamber; many appear as a sharp dot. This type of track is readily distinguished from the extraneous background. Figure 4 is a strip of motion picture film taken of a diffusion chamber with the tritium concentration (T/H ratio) of (1.4×10^{-18}). The arrows point to the tritium tracks. The appearance of the tritium tracks was ascertained by passing some tank deuterium through a Pd thimble and injecting it into the chamber.

Although tritium can be recorded and counted by flash pictures or by movies, the simplest and most sensitive method is simply to visually observe the tracks as they appear. Different observers without previous training agreed in their counting as well as the statistical error permitted. With a little practice observers can count the tritium tracks occurring within a 6-in. circle. This fact determined the size of the diffusion chamber.

A large part of the data was taken by two independent observers, each without knowledge of the origin of the sample or of the others results.

RESULTS

Table I presents the data. Each measurement is the result of 10 to 20 minutes of separate visual counting by two independent observers. The tritium-like tracks are counted both through an eyepiece limiting the view to a $2\frac{1}{2}$-in. circle and with unlimited view of the whole chamber. The background through the eyepiece is 0.70±0.06 track/min; in the whole chamber it is 3.26±0.14 tracks/min.

The diffusion chamber is calibrated both with tritiated hydrogen gas and with tritiated water. The tritiated hydrogen gas is tank deuterium ($T/H=3.6\times10^{-12}$) diluted with tank hydrogen to the concentration, ($T/H=3.6\times10^{-16}$). This filling gives 2.50±0.20 counts/min through the eyepiece, and 11.2±0.5 counts/min in the whole chamber. Tritiated water assayed to have a tritium concentration of 9.0×10^{-12} is diluted with New Jersey well water in the ratios of 1 part per 10 000; 1 part per 50 000; 1 part per 250 000, and 1 part per 1.5×10^{6} parts of New Jersey well water. Since the New Jersey well water itself was measured to have natural tritium concentration of 3×10^{-18}, its natural tritium raises the most dilute calibration sample from 6×10^{-18} to 9×10^{-18}. The results obtained with these calibrated waters are given in Table I.

The results that have been obtained with various natural waters are also given in Table I. One of the interesting features of the results is the variation of the tritium concentration between different samples, even between different rains there are great differences. The May 26th rain over Long Island had a T/H ratio of 800×10^{-18}, while the August 17th rain over Long

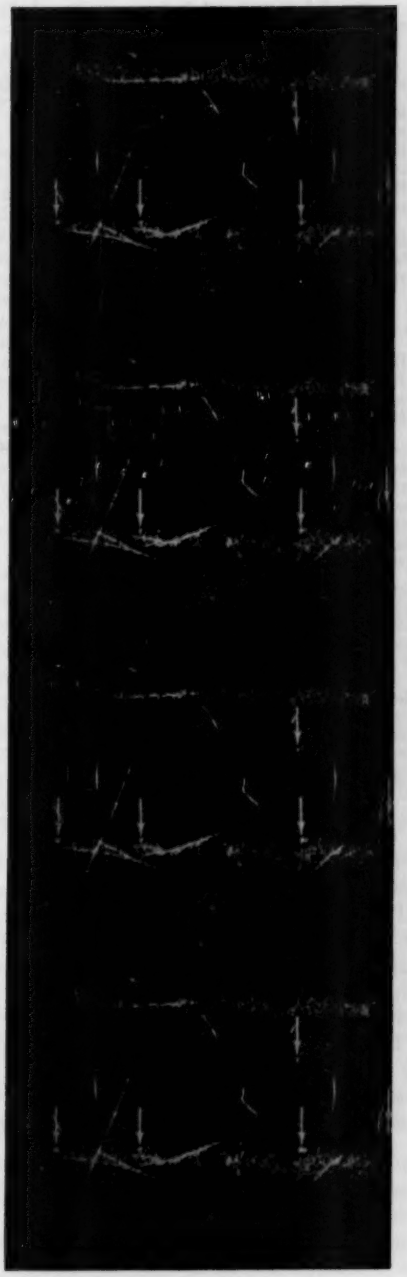

FIG. 4. Strip of 16-mm motion picture film of chamber. (Arrows point to tritium tracks.)

TABLE I. Summary of data.

Sample	View through eyepiece: Tracks/min	View through eyepiece: Tracks-background/min	Enrichment factor	T/H×10¹⁸	Whole chamber view: Tracks/min	Whole chamber view: Tracks-background/min	T/H×10¹⁸
8/2/53 Rain	2.45±0.30	1.75±0.30	1	350±60	8.2±0.8	5.0±1.0	225±45
8/2/53 Rain	2.08±0.23	1.38±0.29	1	276±58	9.6±0.8	6.4±1.0	288±45
8/2/53 Rain (electrolyzed)	20.5±1.0	19.8±1.1	14	282±16	100±6	97±6	311±19
8/2/53 Rain (electrolyzed)	3.55±0.30	2.85±0.36	2	285±36	16.90±0.65	13.74±0.8	310±18
Long Island Sound H_2O	1.05±0.23	0.35±0.29	1	70±60	3.6±0.8	0.34±1.0	15±45
Long Island Sound H_2O (electrolyzed)	0.98±0.16	0.28±0.21	10	5.6±4.0	5.43±0.38	2.17±0.50	9.8±2.3
5/26/53 Rain	4.2±0.3	3.5±0.3	1	700±70	18.2±0.7	15.0±0.8	800±50
7/30/53 Rain	3.68±0.30	1.98±0.40	13	30±10	17.6±1.2	14.3±1.3	49±10
Glacial H_2O	0.78±0.14	0.08±0.20	1	<40	3.48±0.28	0.22±0.42	<20
Glacial H_2O (electrolyzed)	0.71±0.13	0.01±0.19	40	<1	3.20±0.27	−0.12±0.40	<0.5
Glacial H_2O	0.63±0.13	−0.07±0.19	1	<40	2.78±0.26	−0.48±0.40	<20
New Jersey Well H_2O	0.75±0.14	0.05±0.20	1	<40	3.50±0.30	0.22±0.44	<20
New Jersey Well H_2O	0.55±0.12	−0.15±0.18	1	<40	3.28±0.29	0.02±0.43	<20
New Jersey Well H_2O (electrolyzed)	0.80±0.14	0.10±0.20	8	<6	4.05±0.33	0.79±0.47	4.5±2.7
New Jersey Well H_2O	0.78±0.14	0.08±0.20	1	<40	3.65±0.30	0.39±0.44	<20
New Jersey Well H_2O (electrolyzed)	0.98±0.15	0.28±0.21	16	3.5±2.6	4.20±0.33	0.94±0.47	2.7±1.4
11/27/53 Moisture	1.05±0.16	0.35±0.22	1	70±45	4.25±0.33	0.99±0.47	45±21
11/24/53 Moisture	0.65±0.13	−0.05±0.19	1	<40	2.88±0.27	−0.38±0.40	<20
10/12/53 Rain	1.23±0.18	0.53±0.24	2.5	40±20	5.55±0.38	2.29±0.52	41±10
8/17/53 Rain	1.85±0.20	1.05±0.26	13	16±5	8.08±0.45	4.82±0.59	16.9±4
H^3 Calibration	1.60±0.20	0.90±0.26	1	180	7.36±0.43	4.14±0.49	180
H^3 Calibration	5.45±0.38	4.75±0.44	1	900	20.1±0.7	16.9±0.8	900
H^3 Calibration (electrolyzed)	2.28±0.24	1.58±0.30	40	9	10.78±0.53	7.52±0.65	9
H^3 Calibration (electrolyzed)	2.40±0.25	1.70±0.31	8	39	9.4±0.5	6.0±0.6	39

Island had a T/H ratio of 16×10^{-18}. Some particular rains may be affected by bomb blasts; however, the majority of rains and the world reservoir of natural tritium will not be affected by such small-scale events. The variation of tritium in most rains is dependent on the origin and recent history of the rain water. Rain consisting of water evaporated from the continents, or water that has spent much time at a high altitude should have appreciably more tritium than rains composed of recently evaporated ocean water. A discussion of the tritium production in the atmosphere and its relation to the world reservoir and distribution of natural tritium has been given by one of the authors.[3]

Water from the Mendenhall glacier[5] was the only water in which the tritium could not be detected with a factor of 40 electrolytic enrichment, indicating a tritium concentration of less than 5×10^{-19}.

The authors would like to thank Dr. Ralph Shutt for several helpful discussions on diffusion cloud chambers.

[5] The authors wish to thank Dr. Baden of the Snow, Ice, and Permafrost Research Establishment for this sample.

Measurement of the Neutron-Proton Cross Section at 1.0 and 2.5 Mev*

R. E. Fields,† R. L. Becker,‡ and R. K. Adair§
University of Wisconsin, Madison, Wisconsin

(Received January 6, 1954)

The total cross section of hydrogen for 1.0- and 2.5-Mev neutrons has been determined by measuring the neutron transmission of samples of 2-2-4 trimethylpentane and graphite. Electrostatically analyzed protons from an electrostatic generator were used to bombard thin targets of tritium absorbed in zirconium to produce 2.5-Mev neutrons, and thin targets of Li_2O to produce 1.0-Mev neutrons. A gas-filled recoil counter served as neutron detector. For both cross-section determinations the geometry of the measurement was such that neutrons scattered from the samples at angles greater than 4.2° in the laboratory system were not detected. A pulse-height discriminator, biased to reject pulses from low-energy neutrons, reduced the background from room-scattered neutrons to less than one percent and eliminated effects of low-energy neutron groups from the $Li^7(p,n)$ reaction and the $O^{18}(p,n)$ reaction. Neutron energy spreads, primarily caused by the finite target thicknesses, were determined by measuring the widths of narrow neutron-scattering resonances in sulfur and carbon. The results of the cross-section measurements are: $\sigma=2.525\pm0.009\times10^{-24}$ cm² at a neutron energy of 2.540 Mev, and $\sigma=4.228\pm0.018\times10^{-24}$ cm² at a neutron energy of 1.005 Mev. These values, together with the values of a_t, a_s, and ρ_t given by Burgy, Ringo, and Hughes yield values of the singlet effective range in the shape independent approximation of $2.48\pm0.20\times10^{-13}$ cm from the 2.5-Mev measurement, and $2.56\pm0.25\times10^{-13}$ cm from the 1.0-Mev observations. These effective ranges are consistent with the proton-proton scattering data and the hypothesis of the charge independence of nuclear forces.

I. INTRODUCTION

THERE is limited evidence establishing the charge independence of nuclear forces. While the similarity of energy levels of mirror nuclei[1] strongly suggests the equivalence of neutron-neutron and proton-proton forces, there is not yet as much information on the equality of neutron-proton forces and the forces between like nucleons. Recent experiments by Bockelman *et al.*[2] have served to determine the states of isotopic spin one in N^{14} and in B^{10} which corresponds to the isotopic spin-one ground states of O^{14} and C^{14}, and to Be^{10} and C^{10}, respectively. If nuclear forces are charge independent these states will form charge triplets, split in energy only by Coulomb forces. Since the energy splitting of these states can apparently be accounted for by Coulomb forces, it seems improbable that nuclear forces are strongly charge dependent. An analysis of the effect of charge independence of nuclear forces on the electromagnetic transitions in light nuclei indicates also that charge-dependent forces are not important.[3,4] While the high-energy n-p and p-p scattering suggests a charge-dependent interaction,[5] the cross sections are not in contradiction with any rules based solely on charge independence.[6]

Information derived from high-energy scattering and from states of light nuclei concern the nucleon-nucleon interaction in states of various relative angular momenta. While this is advantageous inasmuch as velocity-dependent forces are effective only in states of angular momentum greater than zero, the complexity of the interaction makes a detailed analysis of the data difficult. At low energies the nucleon-nucleon interaction takes place almost entirely in the S state. Recent theoretical investigations using the effective range theory of Schwinger[7-9] have emphasized the usefulness of low-energy measurements on the n-p and p-p system in the determination of the possible charge dependence of the nucleon-nucleon potential. The effective-range theory shows that, up to energies of a few Mev, the nucleon-nucleon interaction for each spin state can be expressed in terms of two parameters, the scattering length a, and the effective range ρ. Charge independence then implies that the singlet n-p scattering length and effective range, a_s and ρ_s, should be equal to the p-p scattering length and effective range. Schwinger[10] has shown that the difference observed between the n-p scattering length and the p-p scattering length is to be expected because of the different magnetic moments of the neutron and proton. Since the singlet state of the nucleon-nucleon system is almost at resonance near zero energy, the scattering length is very sensitive to variations of the potential and the approximate equality of the scattering lengths suggests

* Work supported by the U. S. Atomic Energy Commission and the Wisconsin Alumni Research Foundation.

† Now at Consolidated Vultee Aircraft Corporation, Fort Worth, Texas.

‡ National Science Foundation Predoctoral Fellow.

§ Now at Brookhaven National Laboratory, Upton, Long Island, New York.

[1] F. Ajzenberg and T. Lauritsen, Revs. Modern Phys. **24**, 321 (1952).

[2] Bockelman, Browne, Buechner, and Sperduto, Phys. Rev. **92**, 665 (1953).

[3] D. H. Wilkinson, Nature **172**, 576 (1953).

[4] L. A. Radicati, Phys. Rev. **87**, 521 (1952).

[5] R. S. Christian and H. P. Noyes, Phys. Rev. **79**, 85 (1950).

[6] D. Feldman, Phys. Rev. **89**, 1159 (1953).

[7] H. A. Bethe, Phys. Rev. **76**, 38 (1949).

[8] J. M. Blatt and J. D. Jackson, Phys. Rev. **76**, 18 (1949).

[9] E. E. Salpeter, Phys. Rev. **82**, 60 (1951).

[10] J. Schwinger, Phys. Rev. **78**, 135 (1950).

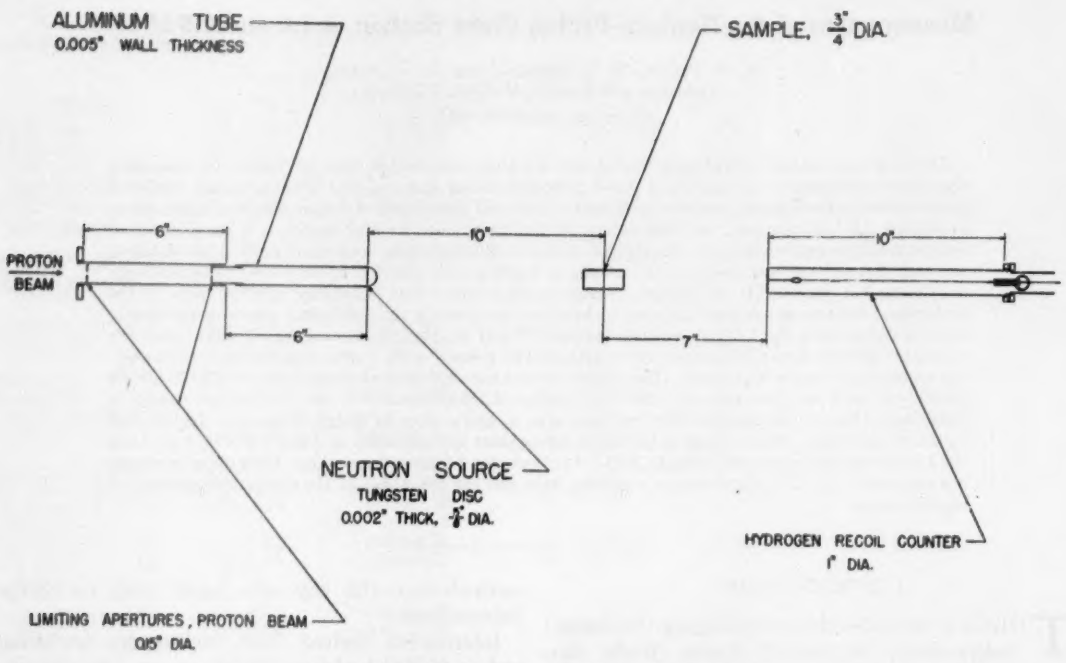

FIG. 1. Schematic diagram of experimental arrangement.

strongly that the n-n and p-p forces are similar. It is, however, possible for two significantly different potentials to give the same scattering length and it is desirable to investigate whether the p-p and n-p effective ranges are the same.

The neutron-proton scattering cross section at low energies is made up of the sum of the triplet and singlet cross sections. The triplet contribution[11] can be calculated from accurately known values of a_t and ρ_t and subtracted from the measured total cross section to determine the singlet scattering which, at low energies, is a function of a_s and ρ_s only. Since a_s is known, ρ_s can then be determined. Salpeter has shown that previous low-energy measurements by Lampi *et al.*[12] indicate a singlet effective range of $2.6\pm0.5\times10^{-13}$ cm in the shape-independent approximation.

Recent accurate n-p cross-section measurements have been made at 5 Mev,[13] at 14 Mev,[14,15] and at 20 Mev.[16] At these energies the cross section depends sensibly on the shape of the potential, therefore the singlet range depends upon the assumptions made concerning the singlet and the triplet well shapes. Conversely, these experiments might be expected to provide information on the shape of the well if the effective range is determined from experiments at lower energies. However, since $(v/c)^2$ is of the order of 3 or 4 percent at neutron energies of 15 or 20 Mev the uncertainties involved in the use of classical dynamics in the interpretation of these measurements may invalidate any very detailed conclusions.

The present experiments were performed to determine the shape-independent n-p effective range in an attempt to obtain further information on the charge dependence of nuclear forces. It was decided to perform cross-section measurements at 1.0 and 2.5 Mev. The difference in energy is sufficiently great that somewhat different techniques of measurement are necessary and the range derived from each measurement will be to some extent independent. This allows a check on the general reliability of the range determination. The effect of shape factors on the cross section at 1.0 Mev should be small, and at 2.5 Mev the shape correction will still be less than the experimental uncertainties.

II. APPARATUS

The measurement of the hydrogen cross section was made, in the case of both the 1.0-Mev and the 2.5-Mev measurements, by determining the relative neutron transmission of a hydrocarbon and of graphite. Figure 1 shows schematically the experimental arrangement. 1.0-Mev neutrons were produced by bombarding Li_2O targets of about 10-kev stopping power with electrostatically analyzed protons accelerated by an electrostatic generator. A target of tritium absorbed in a

[11] Burgy, Ringo, and Hughes, Phys. Rev. **84**, 1160 (1951).
[12] Lampi, Freier, and Williams, Phys. Rev. **80**, 853 (1950).
[13] Hafner, Hornyak, Falk, Snow, and Coor, Phys. Rev. **89**, 204 (1953).
[14] Poss, Salant, Snow, and Yuan, Phys. Rev. **87**, 11 (1952).
[15] Coon, Bondelid, and Phillips, quoted by Coon, Graves, and Barschall, Phys. Rev. **88**, 562 (1952).
[16] R. B. Day and R. L. Henkel, Phys. Rev. **92**, 358 (1953).

30-kev thick layer of zirconium was used to produce 2.5-Mev neutrons.

An important consideration in the design of the neutron source was to reduce to a minimum the number of low-energy neutrons produced by scattering in the source assembly. Most of these neutrons are produced by the primary Li(p,n) or T(p,n) reaction at angles other than 0° with respect to the incident proton beam, and are then scattered elastically in the forward direction by material in the immediate neighborhood of the source. Such neutrons are not counted as background because they are scattered from material near the primary source and cannot be distinguished from the primary beam. Since these scattered neutrons are produced initially at an angle with the proton beam, they have a lower energy than those emitted in the forward direction and therefore add a low energy component to the beam. In order to minimize the material near the source the targets were backed by a wolfram disk 0.002 in. thick. The disk was held in a vacuum housing of 0.005-in. aluminum.

The proton beam passed through two 2.4-mm collimation holes, spaced 15 cm apart, before striking the target. The target and vacuum housing, which are insulated from the collimation system by a length of Lucite, form a Faraday cage. The charge incident on the target was measured by a current integrator. Currents of about one microampere were used throughout both measurements. Since the energy developed in the targets could be dissipated only by radiation, the targets attained temperatures estimated to be about 400°C. It was, therefore, not possible to use lithium, which melts at about 184°C, as a target material. Targets of Li_2O prepared by allowing evaporated lithium to oxidize, proved stable under bombardment. The tritium-zirconium target, used for the 2.5-Mev determination, lost about 20 percent of its tritium in the 200 hours during which it was subjected to bombardment.

A high-pressure gas recoil counter was chosen as a detector, because it had been found in previous work that counters of this type have a high neutron sensitivity, and discriminate effectively against gamma rays. The counter consisted of a stainless-steel outer cylinder, one in. in diameter and 0.030 in. thick, with a 0.040-in. end cap. A 0.005-in. Kovar wire, entering the counter through a Kovar seal, was held at a positive potential. For the detection of 2.5-Mev neutrons, the counter was filled with 35 atmos of tank hydrogen, purified by passing it over palladized asbestos, through a liquid air trap, and over hot magnesium chips. The counter was operated as a pulse ionization chamber with the counter wire held at about 3500 volts. Pulses from the neutron recoils seemed to be somewhat reduced in size compared to those measured at lower gas pressures, possibly because of recombination.

The efficiency of the recoil counter as a function of energy was measured by comparing the counting rates of the recoil counter with those of a long counter. The absolute sensitivity of the long counter was measured by determining the counting rate produced by neutrons from a calibrated radium-beryllium source. Since the electric field in the recoil counter varies considerably from the center wire to the counter wall, it seemed probable that recoils occurring in low field regions near the wall would be affected to a greater extent by recombination and produce a smaller voltage pulse than recoils occurring in high field regions near the wire. If such an effect is important, the efficiency of the counter will be smaller than the efficiency calculated from the known n-p cross section. The circles in Fig. 2 show the measured counter efficiency with the discriminator bias adjusted to the level used in the 2.5-Mev transmission measurements. The solid line represents the calculated efficiency neglecting end effects and the scattering of neutrons into and out of the active volume of the chamber by the end cap and counter walls. In view of the uncertainty of such effects the agreement of the measured and calculated values is satisfactory, indicating that there is no appreciable effect from recombination.

For the 1.0-Mev experiment it appeared desirable to use some gas multiplication to improve the signal-to-noise ratio. Satisfactory operation was obtained by filling the counter with 25 atmos of helium and 0.2 atmos of CO_2. A gas multiplication of about 5 was produced by operating at a voltage of about 3000. The efficiency of the counter is higher than that of a counter filled to the same pressure of hydrogen because of the large back-scattering cross section of helium for 1.0-Mev neutrons. The solid circles in Fig. 2 show the counter efficiency as a function of energy with the same discriminator bias as was used in the 1.0-Mev transmission measurements.

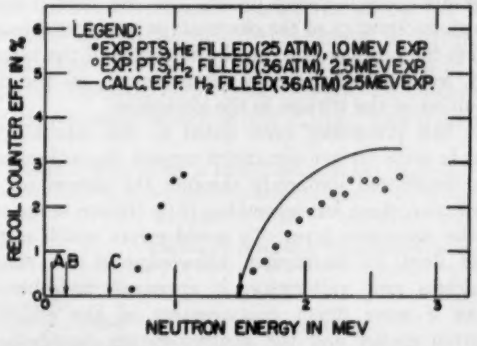

FIG. 2. Counter efficiencies. The solid points show the measured efficiency of the counter as used at 2.5 Mev and the solid curve is the theoretical efficiency. The efficiency of the counter as used at 1.0 Mev is shown by the circles. The three arrows A, B, and C refer to the measurements at 1.0 Mev. Arrows A and B represent the two groups of neutrons at 0° in the laboratory system from the $O^{18}(p,n)F^{18}$ reaction, while arrow C represents the neutrons from the reaction $Li^7(p,n)Be^{7*}$.

If the ionization produced in the counter sets up sufficient space charge to affect the gas multiplication, the counter efficiency might be expected to vary with counting rate. In an effort to determine whether such an effect was important, the neutron counting rate was measured when a 2-mg radium source was placed 10 cm from the counter. Through the radium γ rays produced a very high ionization density in the counter, the neutron counting rate was not affected.

During the 2.5-Mev measurement the neutron flux was monitored by a set of ten $B^{10}F_3$ proportional counters set in a block of paraffin. This monitor was placed 65 cm from the neutron source at an angle of 120° with the proton beam. It was observed in the course of the experiment that the monitor counting rate was accurately proportional to the proton current on the target. This showed that the effects of beam and target nonuniformity were very small, so that for the 1.0-Mev cross-section measurement the flux was monitored by measuring the charge incident on the target by means of a current integrator. Small fluctuations in the proportionality of neutron flux to incident proton current would be randomly distributed among carbon and hydrocarbon transmission measurements. Therefore, the only effect of such fluctuations on the data would be a greater dispersion of the cross-section values measured in individual runs than that indicated by statistical considerations.

III. ENERGY MEASUREMENT

Since the n-p cross section varies approximately as the square root of the neutron energy at the energies at which the cross-section measurements were made, it was necessary to determine the effective neutron energy accurately. For a uniform target the effective neutron energy and the spread in neutron energy can be calculated from the energy distribution of the incident protons, since the reaction energy is known. The effective neutron energy distribution will depend upon the characteristics of the electrostatic analyzer through which the protons pass, on the thickness of the target, and, for the higher-energy measurement, on the distribution of the tritium in the zirconium.

It had previously been noted at this laboratory[17] that in some tritium-zirconium targets the tritium was not distributed uniformly through the zirconium. In particular, there was sometimes little tritium in the top of the zirconium layer. To avoid errors which might arise from an incomplete knowledge of the target thickness and uniformity, it appeared desirable to make a more direct measurement of the effective neutron energy and the neutron-energy distribution. It is possible to determine the shape of the neutron-energy distribution by measuring the yield from a neutron-induced reaction at a very narrow resonance as a function of the incident-proton energy. By increasing the proton energy in small steps, ΔE_p, the neutron-energy spectrum can be moved past the resonance in small steps, ΔE_n, where $\Delta E_n \approx \Delta E_p$ if E_p is far enough above threshold. The resonance, if it is narrow compared to the width of the neutron-energy distribution, may be regarded as a detector which picks out of the spectrum neutrons with a particular energy. Therefore a measurement of the yield as a function of E_p enables one to determine the neutron spectrum.

Bockelman *et al.*[18] have found a resonance, about 8 kev wide, for the scattering of 2.08-Mev neutrons by carbon which could be used to determine the mean energy and the energy spread of the neutrons from the tritium target. An accurate measurement was made of this resonance energy so the resonance could be used as a neutron-energy standard. A thin lithium target was evaporated in place and used to measure the transmission of carbon at this resonance. Resolution slits on the electrostatic analyzer were adjusted so that the proton energy half-width was about 2 kev.

Values of the transmission are shown by solid circles in Fig. 3. The abscissas of the figure represent potentiometer readings which are proportional to the voltage across the analyzer plates and hence to the proton energy. Crosses on Fig. 3 represent the counting rate as a function of proton energy at the Li(p,n) threshold using the H_2^+ beam from the generator. This threshold measurement establishes the absolute energy scale for the 2.5-Mev measurement. Excepting small corrections for relativistic effects and the mass of the electron in the H_2^+ ion, the energy scale represented by the abscissa of Fig. 3 for the threshold measurement is just one-half of that for the transmission measurement. A nonlinearity of the analyzer would not affect the energy measurement as the

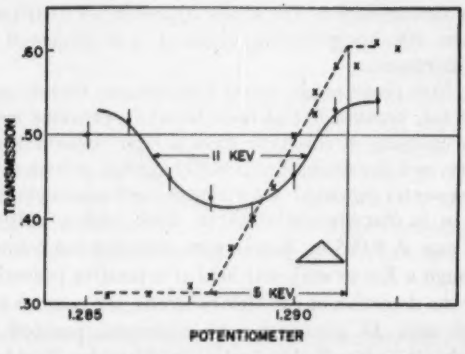

FIG. 3. Energy of carbon resonance. The broken line (crosses) represents the threshold rise curve of neutron counting rate, obtained by bombarding a thin lithium target with molecular hydrogen ions. The solid line (solid circles) shows the transmission of a carbon sample measured with neutrons obtained by bombarding the lithium target with protons. The abscissa represents a potentiometer reading which is proportional to the voltage across the plates of the electrostatic analyzer.

[17] Williamson, Browne, Craig, and Donahue, Phys. Rev. **84**, 731 (1951).

[18] Bockelman, Miller, Adair, and Barschall, Phys. Rev. **84**, 69 (1951).

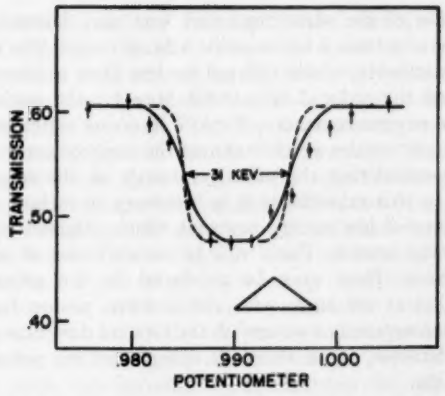

FIG. 4. Carbon resonance observed with the tritium target. The experimental points represent the transmission of a carbon sample measured with neutrons obtained by bombarding the tritium target with protons. The solid curve shows the transmission calculated if the target were uniform and 31 kev thick, while the dotted curve shows the transmission shape expected if the resonance were very narrow, and hence is a measure of the neutron-energy spectrum.

analyzer voltage is almost exactly the same for the transmission measurement and for the threshold determination. However, a measurement of the Li(p,n) threshold using the proton beam was in agreement with previous measurements establishing the linearity of the analyzer system. Taking the Li(p,n) threshold as 1.8815±0.001 Mev,[19] the carbon-scattering resonance occurs at a neutron energy of 2.087±0.0035 Mev. The stated error arises from an uncertainty of 2 kev for the value of the center of the resonance, of 2 kev in the energy scale, and about one kev for the half-thickness of the lithium target. The target thickness was estimated as 5 kev for 2-Mev protons from the measured rise of counting rate at threshold.

Results of a transmission measurement at the same resonance using the tritium target are shown in Fig. 4. The solid circles represent the measurements. If the width of the resonance were negligible, the deviation of the points from the constant background would be nearly proportional to the actual neutron-energy distribution. Tritium evenly distributed in a zirconium layer 31 kev thick for protons of this energy would produce a transmission curve shown by the solid line. This curve, which includes the effects of the finite width of the resonance level, appears to be in fair agreement with the experimental points. The dashed curve shows the neutron-energy distribution from such tritium target.

From the kinematics of the T(p,n) reaction[20] it is possible to calculate the proton energy required to produce 2.087-Mev neutrons. This will be the energy of protons at the half-thickness of the zirconium-tritium target. Comparison of this value with the incident proton energy determined from the voltage on the plates of the electrostatic analyzer indicated that there was no appreciable layer of either carbon or nontritiated zirconium at the surface of the target.

The neutron-proton cross section was measured at a neutron energy about 450 kev higher than the carbon-scattering resonance. From the T(p,n) kinematics the increase in proton energy required to reach this neutron energy was calculated. As the increase in proton energy was not large, the change in stopping power of the zirconium was small and the neutron-energy spectrum should be similar to the spectrum determined at 2.08 Mev shown by the dashed line in Fig. 4. Assignment of the effective neutron energy was made from the measured neutron energy at the carbon resonance and the knowledge of the increase in proton energy.

It was concluded that the mean energy of the neutrons used in the cross-section measurement was 2.540±0.006 Mev and the shape of the distribution was essentially that shown by the dashed line in Fig. 4 with a full width of about 30 kev. The error is primarily the result of uncertainties in determination of the center of the distribution, and the error assigned to the energy of the carbon resonance. Table I shows an estimate of all uncertainties affecting the assignment of the error in the energy determination.

At 1.0 Mev the Li(p,n) reaction appeared to be more suitable as a neutron source than the T(p,n) reaction as it was possible to prepare lithium targets which gave a greater neutron yield per kev stopping power, and the low-energy group of neutrons from this reaction could be effectively biased out. It is probable that in the preparation of the Li_2O targets which were used in these experiments some Li_2CO_3 and LiOH were formed. Originally the targets were light gray in color, but they turned white after a brief bombardment by the proton beam. It is believed that most of the water and carbon dioxide was driven off by the heat produced by the beam. The thickness of the Li_2O targets was measured by determining the rise of the yield at threshold and, more reliably, by measuring the experimental width of the 1.5-kev wide neutron-

TABLE I. Uncertainty in energy measurement (in kev).

	2.1-Mev carbon resonance	1-Mev n-p measurement	2.5-Mev n-p measurement	1.3-Mev oxygen resonance
Energy of Li(p,n) threshold	2	1		1
Energy of 2.1-Mev carbon resonance			3.5	
Target thickness	1.5	1.5	3	1.5
Setting for Li(p,n) threshold	2	2		2
Setting for 2.1-Mev carbon resonance			3	
Total (rms)	3.5	3	5.5	3

[19] Jones, McEllistrem, Douglas, and Richards, Phys. Rev. **91**, 482 (1953).

[20] Taschek, Argo, Hemmendinger, and Jarvis, Phys. Rev. **76**, 325 (1949).

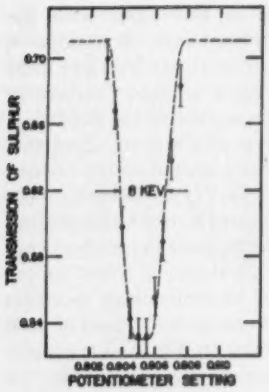

FIG. 5. Lithium oxide target thickness. The curve represents the variation of the transmission of sulfur expected from a target 8 kev thick, the points represent the measured transmissions.

scattering resonance in sulfur at 587 kev.[21] The circles in Fig. 5 show the measured transmission at this resonance with one of the two targets used in the 1.0-Mev cross-section measurement, and the dashed curve shows the transmission expected from a target 8 kev thick with allowance for the spread in energy of the proton beam and the width of the resonance. The measured transmission appears to show an asymmetry indicating a possible nonuniformity of the target thickness. An examination of the Li_2O surface under a microscope supports the view that the surface is sufficiently roughened by the formation of crystalline grains to produce such an effect. This asymmetry is not sufficiently large to affect the measurement adversely.

The n-p cross-section measurement was made at 1.005±0.003 Mev. This energy was calculated from the kinematics of the Li(p,n) reaction and the measured proton energy. Most of the error assigned to the energy is due to the uncertainties in the target thickness and in the stability of the voltage supply of the electrostatic analyzer. Estimates of the contributions of these errors are made in Table I. Electrostatic analyzer dynamics and reaction kinematics were calculated relativistically throughout. Storrs and Frisch[22] have measured the n-p cross section at a neutron energy of about 1.31 Mev. They used a scattering resonance in oxygen at this energy as a reference point and determined its energy to be 1.314 Mev. It appeared desirable to make an independent measurement of this energy to relate more clearly the present measurements to those of Storrs and Frisch, and to establish another point for comparison of neutron-energy measurements. The circles in Fig. 6 are the results of measurements of the transmission of a BeO sample at this resonance using as a neutron source the same Li_2O target that was used for most of the 1-Mev n-p cross-section experiment. The location of the same resonance was also determined by use of a fresh 3-kev metallic lithium target. The two measurements, which differed by less than a kilovolt, yielded the value 1.315±0.003 Mev for the position of the oxygen resonance. Table I shows an estimate of the uncertainties which make up the assigned error.

In considering the effective energy of the sources used in this experiment it is necessary to include the presence of low energy neutrons which originate very near the source. There will be two sources of such neutrons. They may be produced in the primary reaction at an angle with the incident proton beam and subsequently scattered in the forward direction into the detector, or in reactions other than the primary reaction.

Two groups of low-energy neutrons produced by extraneous reactions result from the bombardment of Li_2O by protons. Arrows on Fig. 2 show the energy of neutrons produced by the $Li^7(p,n)Be^{7*}$ reaction and the $O^{18}(p,n)F^{18}$ reaction. Pulses from neutrons of both these reactions will be rejected by the pulse-height discriminator. Though no (p,n) reaction of a stable nucleus can produce neutrons of appreciably higher energy than the T(p,n) reaction, it appears possible that low-energy neutrons might be produced in small numbers by the bombardment of zirconium or wolfram by protons. A measurement of the intensity of such neutrons was made by bombarding a target of zirconium of about 100-kev stopping power evaporated onto a wolfram backing with 3.5-Mev protons and observing the neutron flux with the recoil counter under conditions similar to those used to detect the T(p,n) neutrons. The counting rate produced in this way was less than 0.1 percent of that caused by the T(p,n) neutrons used in the cross-section measurement.

Neutrons produced by the T(p,n) reaction at angles greater than about 70° with respect to the proton beam, or by the Li(p,n) reaction at angles greater than about 80°, have too small an energy to be registered by the counter. However, those emitted at smaller angles but with energies large enough to be counted, can be scattered into the forward direction by the wolfram target backing and by the aluminum vacuum housing.

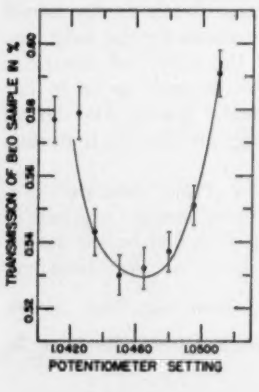

FIG. 6. Energy measurement of 1.31-Mev oxygen resonance. The circles show the transmission of a BeO sample measured with an 8-kev Li_2O target.

[21] Peterson, Barschall, and Bockelman, Phys. Rev. **79**, 593 (1950); Kiehn, Goodman, and Hansen, Phys. Rev. **91**, 66 (1953). The results of Peterson *et al.* made at this laboratory, give a resonance energy of 587 kev, recalculated using relativistic dynamics. Kiehn *et al.* quote a value of 588 kev.

[22] C. L. Storrs and D. H. Frisch, Phys. Rev. **90**, 339 (1953) and private communication.

An estimate of the fraction of the measured flux resulting from this effect was made assuming that the differential cross sections for scattering by wolfram and by aluminum are equal to the total cross sections divided by 4π and that the neutrons are emitted isotropically by the source. The scattered flux was integrated over the portions of the target backing and aluminum vacuum housing from which neutrons could be scattered into the detector with sufficient energy to be recorded. This calculation indicated a contribution of about 0.2 percent of lower energy neutrons for the measurement at 1.0 Mev and considerably less for the measurement at 2.5 Mev. When the counter sensitivity and the small variation of the hydrogen cross section are taken into account, it appears that the effect of these neutrons on the measured cross section can be safely neglected.

IV. CROSS-SECTION MEASUREMENT

The cross section was measured by determining the ratio of the transmission of a thin-walled cylindrical aluminum container filled with 2–2–4 trimethylpentane, $(CH_3)_3CCH_2CH(CH_3)_2$, to the transmission of a similar container holding a graphite cylinder which had the same number of carbon nuclei per square cm as the hydrocarbon samples. Under conditions of good geometry the n-p cross section for monoenergetic neutrons is then equal to $-(1/N)\ln T$, where N is the number of hydrogen atoms per cm^2 in the hydrocarbon scatterer and T represents the ratio of the transmission of the hydrocarbon scatterer to that of the carbon scatterer. It is unnecessary to measure separately either the carbon cross section or the attenuation of the beam by the containers.

The trimethylpentane was obtained from Eastman Kodak; the "spectro" grade, which was used, is stated by the supplier to contain less than 0.1 percent impurities. Using a specific gravity bottle, the density of the liquid as a function of temperature was found to be $0.6921[1+0.0013(20°-t)]\pm0.0003$ g/ml, in agreement with the handbook value. Since the most probable impurities are hydrocarbons with carbon-hydrogen ratios similar to trimethylpentane, the impurities will probably introduce a negligible error. For similar reasons the presence of dissolved water cannot affect the validity of the measurement. It was, however, necessary to monitor the temperature of the liquid carefully because of the large thermal expansion coefficient. The temperature of a brass block on which the sample rested during measurements on the carbon sample and of the background was measured to within 0.2°.

The aluminum containers were cylinders of 0.75-in. inside diameter with walls 0.005 in. thick. For the 2.5-Mev measurement the scatterer was about 4.4 cm long and had end caps of aluminum 0.005 in. thick which were waxed into place. The scatterer used for the 1.0-Mev measurement was about 3 cm long and had end caps of 0.005-in. tantalum. The length of the scatterers was measured with a caliper the opening of which was determined using a traveling microscope. The engine screw of the microscope had been calibrated with standard meter bar BS No. 210. Individual determinations of the length of the scatterer were reproducible to within 0.0003 cm. Since the end caps were not perfectly flat or parallel, measurements were made at many points on the end surfaces. The mean deviation of the thickness along the axis was about 0.0020 cm. An average length was computed from these measurements. When the transmission measurements at 2.5 Mev were in progress it was noted that the cylinder dimensions were not completely stable under the necessary handling involved in changing samples. The average length of the sample varied by as much as one part in a thousand from day to day, though the mean deviation was about one-half of this. The greater strength of the tantalum end caps used on the scatterer for the 1.0-Mev measurement resulted in much greater dimensional stability of that sample. Carbon scatterers were machined from spectroscopically pure graphite rods obtained from the National Carbon Company. Density of the graphite was measured to be 1.48 g/cm^3 for several samples, indicating no appreciable variation in density through the cylinders. Since the scatterers did not change weight appreciably when heated, it was concluded that large amounts of absorbed gases were not present.

The scatterers rested in position on a light wire cradle mounted on a thin vertical rod supported by a base on the floor. Alignment was checked periodically by placing on the cradle a Lucite tube, of the same outside diameter as the scatterers and a length equal to the distance from the source to the face of the counter. Deviations of the centers of the source, sample, and counter, of 1 mm from a straight line were easily detected and corrected. Since the number of atoms per cm^2 in the scatterer varies inversely as the cosine of the angle between the axis of the scatterer and the beam direction, it is necessary to align the scatterer accurately. With the aid of the Lucite tube deviations of one degree could be observed. Alignment checks were made continually throughout the measurements.

Background counts, resulting mainly from room-scattered neutrons, were determined by measuring the counting rates with the primary beam of neutrons blocked out by scatterers of paraffin or Lucite. A Lucite cylinder, 35 cm long, having a calculated transmission of 0.001 was used to shield the direct beam to determine the background for the 2.5-Mev measurement. The magnitude of this background was about 0.6 percent of the counting rate from the unshielded direct beam. A paraffin cone 30 cm long with a negligible calculated transmission was used in the same way to show that the background at 1.0 Mev was about 0.8 percent.

Cross sections calculated on the assumption that

attenuation by the scatterer is an exponential function of the length of the scatterer are subject to errors caused by the hardening effect and by neutrons scattered into the detector by the sample. The hardening effect is the result of variation of the cross section over the finite neutron-energy spread. Neutrons on the low-energy side of the spectrum are scattered with slightly higher probability than those on the high side. Therefore the mean energy increases as the beam passes through the scattering sample. The result is a rate of decrease of intensity slightly less rapid than exponential.

Calculation of the cross section from a measured average transmission for finite neutron-energy spread, assuming exponential attenuation, therefore, yields a value less than or equal to the true cross section averaged over the energy spread. The equality holds in the limit of vanishing attenuation or if the cross section is independent of energy. In the present experiments the hydrogen cross section changes about ½ percent over the energy spectrum. The resulting error is only about 0.02 percent.

The error caused by neutrons scattered into the detector by single collisions in the sample is given by

$$\frac{\delta\sigma}{\sigma}=\frac{g}{4}\cdot\left[\frac{r(D_1+D_2)}{D_1D_2}\right]^2, \tag{1}$$

where r is the radius of the scatterer, T is its transmission, D_1 is the distance from the source to the scatterer, and D_2 is the distance from the scatterer to the detector.

$$g=\left(\frac{d\sigma}{d\omega}\right)_{0^\circ}\Big/\frac{\sigma_t}{4\pi},$$

i.e., g is the differential cross section at 0° divided by the average differential cross section averaged over all directions; it is equal to 4 for hydrogen. Values of g for carbon are somewhat uncertain. At 2.5 Mev, $g=4$ was selected on the basis of published measurements of the angular distribution of neutrons scattered from carbon,[23] while at 1.0 Mev, $g=2.5$ was used. In-scattering effects from carbon cancel out to some extent so that the uncertainty in angular distribution has only a small effect on the in-scattering correction applied to the relative transmission.

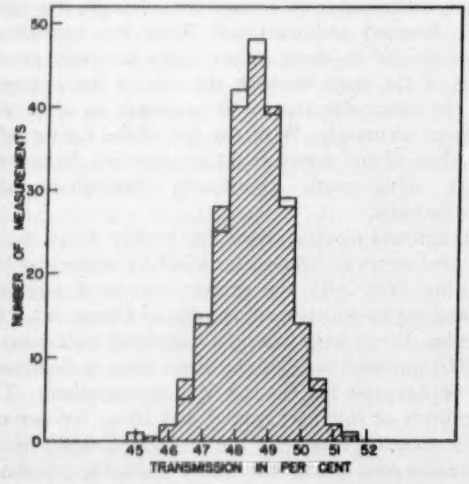

FIG. 7. Histogram of 224 transmission measurements at 2.5 Mev. The frequency is plotted against the transmission. The cross-hatched area represents the "probable" histogram calculated from the number of counts recorded.

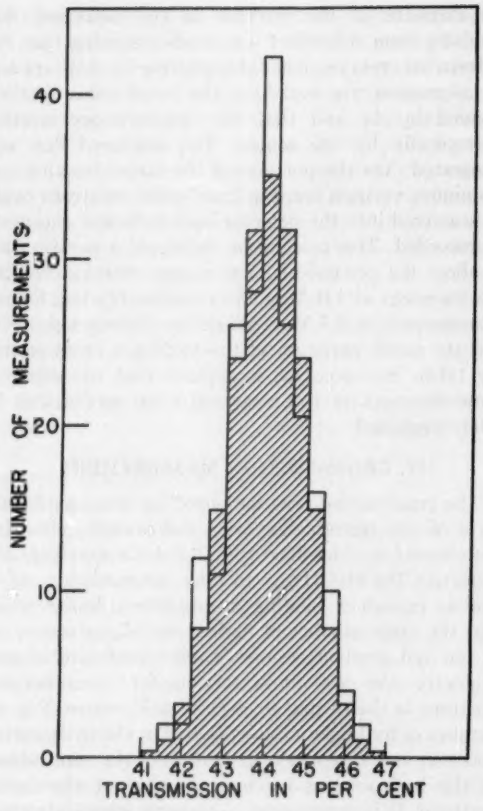

FIG. 8. Histogram of 200 transmission measurements at 1.0 Mev. The frequency is plotted against the transmission. The cross-hatched area represents the "probable" histogram calculated from the number of counts recorded.

The calculated value of $\delta\sigma/\sigma$ is 0.008 for the 2.5-Mev measurement and 0.007 for the 1.0-Mev measurement.

Neutrons reaching the detector after more than one collision in the scatterer constitute a fraction less than $(1-T-T\ln T)gd\omega$ of the direct beam, where $d\omega$ is given by $\frac{1}{4}[r(D_1+D_2)/D_1D_2]^2$. In the present experiments this amounts to about two-thirds of the fraction singly in-scattered. However, if the effect of energy degradation on the detection efficiency for the multiply in-scattered neutrons is taken into account it appears that

[23] Baldinger, Huber, Ricamo, and Zunti, Helv. Phys. Acta **23**, 503 (1950).

the net correction for multiple in-scattering is not more than one-tenth of the correction for single in-scattering.

Transmissions were determined at each energy by making over 200 individual measurements, each of which determined the cross section with a statistical standard error of 2.7 percent. Figures 7 and 8 show the distribution of these measured transmissions, together with normal distributions calculated from the number of counts recorded in each measurement. The standard errors in the averages of the experimental measurements are 0.14 percent at 2.5 Mev and 0.17 percent at 1.0 Mev. The standard errors predicted from the number of counts were in each case almost exactly the same. This good agreement between the errors calculated from the individual measurements and from the number of counts indicates that any random variation of experimental conditions, such as fluctuations in the electronic circuits, does not seriously affect the accuracy of the measurements.

Counting rates were less than 30 counts per second for the 2.5-Mev experiment and 20 counts per second at 1.0-Mev. The amplifier had a rise time of 0.5 μsec, and clipping times of 2 μsec and 4 μsec were used in the 2.5- and 1.0 Mev experiments, respectively. Measurements with a double pulse generator gave a resolution of 5 μsec for the discriminator. The portion of time during which the counter system is dead will be equal to $2Nt$, where N is the number of counts/second and t is the recovery time of the system. The time t will be about equal to the time a pulse stays above the discriminator level plus the recovery time of the discriminator. This recovery time will depend upon the height and shape of the pulse. For the conditions of this experiment the average value of t will be a little less than the sum of the discriminator time constants and the amplifier clipping time constant. Pulses smaller than the setting of the discriminator will be occasionally superposed and register as a count. The fraction of pulses produced by this "pile up" will be equal to $2N't'$, where the average value of t' depends on the shape of the pulses and upon the pulse spectrum, and N' represents the number of pulses per second which are too small to trigger the discriminator. N' is not greatly different from N and the effective value of t' is estimated to be about 2 μsec. The two corrections, for dead time and for pile up, are in opposite directions and have an effect of only about 0.03 percent on the cross section.

TABLE II. Sources of error in cross-section measurement (in parts per thousand).

	2.5-Mev measurement	1.0-Mev measurement
Statistical standard error	1.9	2.3
Energy	1.9	1.5
Background	1.2	1.6
In-scattering	1.3	1.6
Nuclei/cm^2	1.5	1.6
Total (rms)	3.5	4.0

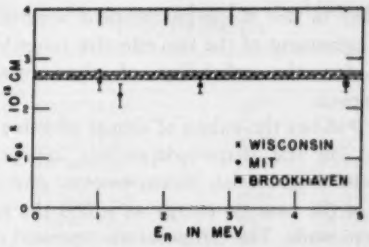

FIG. 9. Singlet effective range vs energy. The points at 1.0 and 2.5 Mev represent effective ranges calculated from the present measurements. The 1.3-Mev and 5-Mev points show ranges calculated from the work of the MIT group and the Brookhaven group, respectively. The errors shown reflect only the uncertainties in the n-p cross-section measurements. The dashed line shows the proton-proton effective range.

V. RESULTS AND DISCUSSION

Evaluation of the two cross-section measurements results in a neutron-proton total cross section of 4.525±0.009 barns at a neutron energy of 2.540 Mev, and a value of 4.228±0.018 barns at 1.005 Mev, where the quoted errors represent standard deviations. Table II lists the important sources of error and estimates of their magnitudes in terms of parts per thousand of the cross section. The uncertainty in neutron energy is treated as an uncertainty in cross section at an exact neutron energy. The resultant total error was taken as the root mean square of the individual errors.

From these values of the cross sections a value of the neutron-proton singlet effective range can be derived and compared with values calculated from neutron-proton cross sections obtained by other workers, and with the proton-proton effective range. Singlet effective ranges were calculated in the shape-independent approximation using the relationship

$$\sigma=\frac{3\pi}{k^2+(1/a_t-\frac{1}{2}\rho_t k^2)^2}+\frac{\pi}{k^2+(1/a_s-\frac{1}{2}\rho_s k^2)^2}, \quad (2)$$

where k is the neutron wave number, a_t and a_s are the triplet and singlet scattering lengths, and ρ_t and ρ_s are the corresponding effective ranges. Values of a_s, a_t, and ρ_t were taken from the paper of Burgy *et al.*[11] At the energies used in this experiment the relativistic corrections to k are small. k^2 was, therefore, calculated from $k^2=2\mu E/\hbar^2$ where E represents the energy of the system in center-of-mass coordinates and μ the reduced mass. The singlet effective range calculated from the 2.5-Mev cross-section measurements is 2.48±0.20 $\times10^{-13}$ cm; while the 1.0-Mev measurement leads to a value of 2.56±0.25$\times10^{-13}$ cm. The errors assigned to the two determinations include uncertainties in the parameters a_s, a_t, and ρ_t as well as the measured cross sections and are, therefore, not independent. Assuming

the validity of the shape-independent approximation the close agreement of the two effective range values is evidence for the reliability of the cross-section measurements.

Figure 9 shows the values of singlet effective ranges, calculated on the shape-independent approximation from recent cross-section measurements, plotted as a function of the neutron energy at which the measurements were made. The errors shown represent only the independent errors in the cross section measurements and do not include the uncertainties in the parameters a_s, a_t, and ρ_t. Figure 9 shows excellent agreement between the various measurements.

The dashed line in Fig. 9 represents the value of the proton-proton effective range[24,25] which is taken as $2.65\pm0.07\cdot10^{-13}$ cm. Though the calculated neutron-proton singlet effective ranges appear to be a little smaller than the proton-proton ranges, the agreement is within the errors of the experiments and constitutes evidence for the charge independence of nuclear forces.

[24] J. D. Jackson and J. M. Blatt, Revs. Modern Phys. 22, 77 (1950).
[25] H. H. Hall and J. L. Powell, Phys. Rev. 90, 912 (1953).

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Atomic Masses in the Intermediate Region*

T. L. COLLINS, WALTER H. JOHNSON, JR., AND ALFRED O. NIER
Department of Physics, University of Minneapolis, Minnesota†

(Received January 7, 1954)

The double-focusing mass spectrometer previously described has been used to measure 28 atomic masses in the region between gallium and niobium. The new information combined with previous determinations permits one to draw a rather complete packing fraction curve for the region from sulfur through xenon. Analysis shows that masses for stable and unstable isotopes of an element can be fitted by parabolas, one for odd A and one for even A. The two parabolas have the same shape (except just above a proton shell) but are displaced in A as well as mass. The displacements are such that, for an isobar, the minimum mass does not occur at the same Z for the four nuclear families (even even, etc.), in contradiction to usual assumptions.

THIS paper presents the results of a portion of an extensive program of mass measurement undertaken with the double-focusing mass spectrometer[1] developed at the University of Minnesota. Previous communications have covered the regions sulfur to zinc[2,3] and palladium to xenon,[4] as well as detailed studies of the triple isobars at mass 40 and 50.[5,6] We now report 28 mass doublets covering isotopes from gallium, $Z=31$, to niobium (columbium), $Z=41$. The combined data is now sufficient to reveal patterns and shell effects which are discussed under the heading of Mass Systematics.

DOUBLET MEASUREMENTS

The method of measurement has not been changed. We measure the atomic mass difference of two ions of the same mass number (a mass doublet). One ion has a known mass—it is usually a hydrocarbon fragment. The ion current is detected electronically, and because of the nature of the feedback arrangement of the control circuits, the measurement of the mass difference reduces to a measurement of a resistance ratio.[3,7] A "run" consists of at least 10 consecutive tracings of the mass spectrum with alternate forward and backward sweep. Runs were taken on different days over a period of months. During these measurements, the width at half height of a mass "peak" corresponded to 1/10 000 of the ion mass. At these masses this resolution is insufficient to separate the C^{13} satellite (one less hydrogen) from the hydrocarbon peak. A correction is possible by measuring the relative intensity of the hydrocarbon ion at the next lower mass number, and where necessary this correction has been applied.

Table I lists the results of the doublet measurements. The errors are probable errors, and are purely statistical. The day-to-day variation usually exceeds that expected from the statistics within the runs. In this case a simple average is computed; otherwise a weighted average is used.

Several stable isotopes are missing. Selenium ions were reliably obtained only from H_2Se gas with the result that Se^{77} and Se^{78} were accompanied by incompletely resolved selenides of other isotopes. Krypton

* Preliminary Report made at Washington meeting of the American Physical Society, April, 1953 [Phys. Rev. 91, 482 (1953)].

† Research supported by the joint program of the U. S. Office of Naval Research and the U. S. Atomic Energy Commission.

[1] E. G. Johnson and A. O. Nier, Phys. Rev. 91, 10 (1953).
[2] Collins, Nier, and Johnson, Phys. Rev. 84, 717 (1951).
[3] Collins, Nier, and Johnson, Phys. Rev. 86, 408 (1952).
[4] R. E. Halsted, Phys. Rev. 88, 666 (1952).
[5] W. H. Johnson, Phys. Rev. 87, 166 (1952).
[6] W. H. Johnson, Phys. Rev. 88, 1213 (1952).
[7] A. O. Nier and T. R. Roberts, Phys. Rev. 81, 507 (1951).

masses have been reported elsewhere;[8] the present results include those measurements and supersede them. Kr^{80} has been omitted. The hydrocarbon fragment contained a large C^{13} satellite, for which it is difficult to correct, and the doublet value is probably in error. Finally, zirconium masses are far from complete.

ATOMIC MASSES

Table II presents the new atomic masses. Consistent with our previous tables, masses are computed using $H^1=1.008146\pm3$, $C^{12}=12.003842\pm4$ from our own measurements.[2] For completeness, masses for Se^{77}, Se^{78}, and Kr^{80} have been added. The masses may be compared with mass differences from reaction and disintegration energies. Table III shows that the comparison is good. Selenium masses may also be compared with the ratio of selenium mass differences available from microwave measurements of Geschwind *et al.*[9] Table IV shows an excellent agreement.

TABLE I. Mass doublets.

Doublet	ΔM (10^{-4} amu)	No. of runs	Source of ions
$C_6H_9-Ga^{69}$	1447.5 ±0.4	5	metal
$C_5H_{11}-Ga^{71}$	1613.0 ±0.8	5	
$C_5H_{10}-Ge^{70}$	1543.0 ±0.6	6	metal
$C_5H_{12}-Ge^{72}$	1723.5 ±0.5	5	
C_6H-Ge^{73}	845.1 ±0.3	5	
$C_6H_2-Ge^{74}$	946.8 ±0.6	7	
$C_6H_4-Ge^{76}$	1100.5 ±0.4	6	
$C_6H_3-As^{75}$	1017.9 ±0.4	7	metal
$C_6H_2-Se^{74}$	931.4 ±0.7	4	H_2Se gas
$C_6H_4-Se^{76}$	1120.6 ±0.4	4	
$C_6H_9-HSe^{80}$	1461.7 ±0.4	4	
$C_6H_{12}-H_2Se^{82}$	1616.6 ±0.4	4	
$C_6H_7-Br^{79}$	1364.2 ±0.5	5	C_2H_5Br
$C_6H_9-Br^{81}$	1540.5 ±0.5	6	
$C_3H_3-Kr^{78}/2$	634.0 ±0.4	5	gas
$C_3H_5-Kr^{80}/2$	824.19±0.28	9	
$C_6H_{11}-Kr^{83}$	1720.7 ±0.5	5	
$C_3H_6-Kr^{84}/2$	912.20±0.26	10	
$C_3H_7-Kr^{86}/2$	994.07±0.32	6	
$C_6H_{13}-Rb^{85}$	1897.5 ±0.6	5	Rb_2SO_4
$C_5H_{11}O-Rb^{87}$	1717.3 ±1.7	6	
$C_6H_{12}-Sr^{84}$	1807.0 ±1.5	8	metal
$C_6H_{14}-Sr^{86}$	2002.5 ±1.0	10	
$C_5H_{11}O-Sr^{87}$	1720.5 ±0.6	6	
$C_4H_9O_2-Sr^{88}$	1464.6 ±1.1	12	
$C_6H_9-Y^{89}O$	1698.4 ±1.1	5	oxide
$C_7H_4-Zr^{90}$	1426.6 ±2.5	4	oxide
$C_7H_9-Nb^{93}$	1648.1 ±0.8	4	metal

TABLE II. Atomic masses based on $H^1=1.008146\pm3$, $C^{12}=12.003842\pm4$.

Element	Atomic mass	Element	Atomic mass
Ga^{69}	68.94778±6	Kr^{78}	77.94513± 9
Ga^{71}	70.94752±9	Kr^{80}	(79.94194± 7)[b]
Ge^{70}	69.94637±7	Kr^{82}	81.93967± 7
Ge^{73}	71.94462±7	Kr^{83}	82.94059± 7
Ge^{73}	72.94669±4	Kr^{84}	83.93836± 7
Ge^{74}	73.94466±6	Kr^{86}	85.93828± 8
Ge^{76}	75.94559±5	Rb^{85}	84.93920± 8
As^{75}	74.94570±5	Rb^{87}	86.93709±17
Se^{74}	73.94620±8	Sr^{84}	83.94011±15
Se^{76}	75.94357±5	Sr^{86}	85.93684±11
Se^{77}	(76.94459±5)[a]	Sr^{87}	86.93677± 8
Se^{78}	(77.94232±5)[a]	Sr^{88}	87.93408±11
Se^{80}	79.94205±5	Y^{89}	88.93421±11
Se^{82}	81.94285±6	Zr^{90}	89.93311±25
Br^{79}	78.94365±6	Nb^{93}	92.93540± 9
Br^{81}	80.94232±6		

[a] Computed from Se^{76} using (n,γ) reaction data of B. B. Kinsey and G. A. Bartholomew, Can. J. Phys. (to be published).
[b] Computed from Se^{80} using beta scheme reported in Hollander, Perlman, and Seaborg, Revs. Modern Phys. **25**, 469 (1953).

MASS SYSTEMATICS

The quantity of precision mass data at intermediate masses is now sufficient to allow an empirical examination for any regularity which might guide the theorist. Because two independent variables (chosen from A, Z, N and T) are required to specify an atom, the description will be in terms of a mass surface. The pattern of stable isotopes shows the necessity of dividing atoms into at least three families: one for odd A and two for even A (even-even and odd-odd). The Bohr-Wheeler formula[10] assumes that these three surfaces are smooth

TABLE III. Comparison with reaction data. Difference between masses expressed as a difference of mass defects $(A-M)$ in milli-mass units.

Atoms	Mass spectrometer difference mMU	Reaction difference mMU	Reactions	Reference
$Ga^{69}-Zn^{70}$	0.01±0.06	0.06±0.2	(γ,n); β	a,b
$Ga^{71}-Zn^{70}$	0.27±0.10	−0.01±0.2	(γ,n); (p,n)	a,d
$Ge^{74}-As^{75}$	1.04±0.07	0.77±0.2	β^+; (γ,n)	b,e
$As^{75}-Se^{74}$	0.50±0.08	0.51±0.2	(γ,n); β	a,b
$Ge^{74}-Se^{74}$	1.54±0.10	1.28	β^+; β	b
$Se^{76}-As^{75}$	2.13±0.06	2.21±0.06	(n,γ); β	a,b
$Se^{76}-Br^{79}$	0.08±0.06	0.24±0.25	(n,γ); (n,γ); (p,n); (γ,n)	a,f,e
$Se^{80}-Br^{81}$	0.27±0.06	0.20±0.14	β^+; (γ,n)	b,e
$Sr^{87}-Rb^{87}$	0.32±0.18	0.30	β	b
$Sr^{87}-Sr^{86}$	0.07±0.12	0.16±0.2	$d\rho$	g
$Sr^{88}-Y^{89}$	0.13±0.15	0.38±0.2	$d\rho$; β	g,b
$Zr^{90}-Y^{89}$	1.10±0.25	1.36±0.3	(p,n); (n,γ)	f,a
$Zr^{90}-Nb^{93}$	2.28±0.25	2.44±0.3	$d\rho$; $(n\gamma)$; (p,n); (γ,n)	g,a,f,e

[a] Hanson, Duffield, Knight, Diven, and Palevsky, Phys. Rev. **76**, 578 (1949).
[b] Hollander, Perlman, and Seaborg, Revs. Modern Phys. **25**, 469 (1953).
[c] Sher, Halpern, and Mann, Phys. Rev. **84**, 387 (1951).
[d] C. C. Trail and C. H. Johnson, Phys. Rev. **91**, 474 (1953).
[e] B. B. Kinsey and G. A. Bartholomew, Can. J. Phys. (to be published).
[f] Blaser, Boehm, Marmier, and Scherrer, Helv. Phys. Acta. **24**, 441 (1951).
[g] J. A. Harvey, Phys. Rev. **81**, 353 (1951).

[8] T. L. Collins, National Bureau of Standards Circular 522 (U. S. Government Printing Office, Washington, D. C., 1953), p. 67.
[9] Geschwind, Minden, and Townes, Phys. Rev. **78**, 174 (1950).
[10] E. Fermi, *Nuclear Physics Notes* (University of Chicago Press, Chicago, 1950), pp. 6–8.

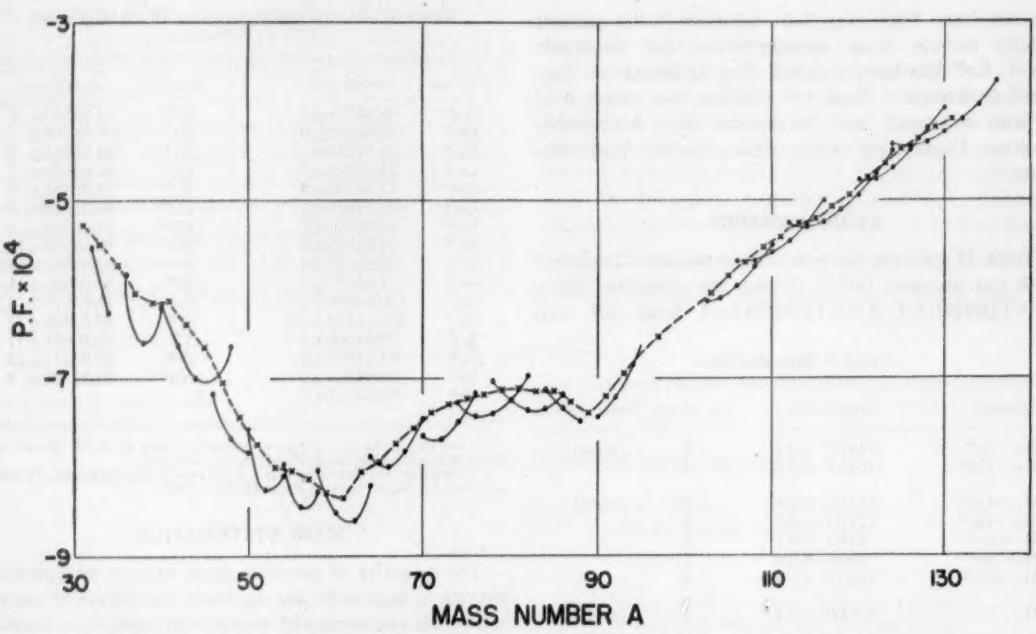

FIG. 1. Packing fraction curve for stable atoms from S^{32} to Xe^{136}. Dashed line is through odd-A atoms; solid lines are through even-A atoms for each element.

and have the same quadratic form for isobars (constant A). The even-A surfaces are assumed to be equally but oppositely displaced from the odd-A surface by an amount of mass which decreases slowly for increasing A. In spite of the fact that numerical constants which have been used in the formula are incorrect, these assumptions have been generally confirmed by beta ray systematics, with two important modifications.[11] First, two surfaces are required for odd A in some mass regions. Second, the surfaces have discontinuities at certain "magic numbers." We can now examine mass surfaces directly. We find much regularity, but the relative position of the three surfaces is considerably more complicated than previously assumed.

For independent variables we shall use A and Z, and for the mass variable we shall use mass defect, $A-M$, in milli-mass units. Imagine that A and Z are horizontal coordinates, and mass defect is downward. Stable atoms are then lower than their unstable neighbors, and the mass surfaces form a narrow valley. Beta-ray data gives information about isobaric (constant A) sections of the valley. We will examine isotopic (constant Z) sections because they contain roughly twice as many points. The minima of the isobaric sections may be joined by a line, representing the greatest stability against beta decay. This line will not coincide with a similar line through the minima of isotopic sections, the difference depending upon the slope of the valley bottom and, hence, upon the mass variable.

A more fundamental variable would be binding energy, which is related to mass defect ΔM by

$$\text{B.E.}=8.986A-0.840Z+\Delta M.$$

Binding energy increases rapidly with A and gives isotopic minima far removed from the region of stability. Our conclusions are not affected by a change to this variable because the transformation is linear. For a plot extending over a wide range, it is common to use $\text{B.E.}/A=8.986-0.840Z/A+\Delta M/A$, or to use the packing fraction, $-\Delta M/A$. The shapes of curves using these variables are almost identical.

Figure 1 gives an over-all picture of our mass measurements. It is a packing fraction curve for stable atoms from S^{32} to Xe^{136}. Errors of the points are about 0.01 unit, too small to show. The relation of the curve to the mass surface is indicated by lines. Odd-A atoms are joined by a single dashed line. There is only one stable atom for each odd A (except 113, 123), so this line is close to the line joining isobaric minima for the odd-A

TABLE IV. Selenium masses from microwave spectroscopy.[a]

	Mass spectroscopy	Microwave
Se^{74}	73.94620±8	73.94659±60
Se^{76}	75.94357±5	(75.94357)[b]
Se^{77}	76.94459±5	76.94456±40
Se^{78}	77.94232±5	77.94219±40
Se^{80}	79.94205±5	(79.94205)[b]
Se^{82}	81.94285±5	81.94275±60

[a] See reference 9.
[b] Two masses assumed in order to compute remaining four masses.

[11] H. E. Suess and J. H. D. Jensen, Arkiv Fysik 3, 577 (1951).

surface. It is apparent that a single line is sufficient for both odd and even Z. Even-A, even-Z atoms are joined by curves for each element. The curves are isotopic sections of the even even surface. The odd-A line shows discontinuities of slope at mass numbers 40 ($20n$,$20p$), 51 ($28n$), 61 ($28p$), 89 ($50n$), and 113 ($50p$). The even even curves accentuate these discontinuities. We suggest that there is a discontinuity near $A=71$ which could be associated with 40 neutrons. Most striking, in Fig. 1, is the second minimum at 50 neutrons.

TABLE V. Mass defect $(A-M)$ for unstable atoms.

Atom	$(A-M)$ mMU	Reaction	Reference
Ni^{57}	40.90±20	(γ,n)	a
Ni^{59}	47.23± 9	(n,γ)	b
Ni^{65}	49.77± 8	β	c
Ga^{66}	46.12± 6	β^+	c
Ga^{67}	50.74± 5	pn	d
Ga^{68}	50.03± 7	β^+	c
Ga^{70}	51.47± 6	pn	d
Ga^{72}	51.08± 6	β	c
Ga^{73}	51.74±20	β	c
Ge^{69}	48.61± 4	β^+	c
Ge^{71}	52.20± 9	pn	d
Ge^{75}	53.08± 6	β	c
Ge^{77}	52.02± 6	β	c
As^{72}	50.70± 6	β^+	c
As^{73}	52.89± 5	pn	d
As^{74}	52.45±10	β^+, β^-, (γ,n)	e,c
As^{76}	53.12± 6	(n,γ); β	f,c
As^{77}	54.66± 6	β	c
Se^{73}	50.01± 6	β	c
Se^{75}	53.34± 4	p,n	d
Se^{79}	56.18± 6	β	c
Se^{81}	56.20± 5	β	c
Br^{75}	50.42± 8	β^+	c
Br^{76}	51.50± 8	β^+	c
Br^{77}	53.95± 8	β^+	c
Br^{78}	53.90±15	(γ,n)	c
Br^{80}	55.92± 7	β^+	c
Br^{82}	58.41± 7	β	c
Br^{84}	56.60± 7	β	c
Br^{86}	53.12±25	β	c
Kr^{79}	54.33± 7	β^+	c
Kr^{85}	60.05± 9	β	c
Kr^{87}	59.01±18	β	c
Sr^{90}	63.90±25	β, β	c
Ag^{111}	59.10±15	β	c
Ag^{112}	57.24±20	β	c
Cd^{107}	59.47±15	β^+	c
Cd^{109}	60.42±15	β^+	c
Cd^{115}	58.10±12	β	c
In^{110}	57.21±14	β^+	c
In^{111}	59.32±25	$(\alpha,2n)$	g
In^{112}	58.19±25	β^+	c
In^{114}	58.23±10	β^+	c
In^{116}	57.56±15	β	c
In^{117}	57.62±15	β	c
In^{119}	55.38±25	β	c
Sn^{113}	59.09±15	β^+	c
Sn^{121}	57.06±15	β	c
Sn^{122}	54.96±20	(γ,n)	h
Te^{127}	53.90±20	β	c
Te^{131}	49.90±42	β	c
I^{124}	53.32±32	β^+	c
I^{125}	55.26±32	β^+	c
I^{126}	53.88±14	β	c
I^{128}	53.37±10	β	c
I^{129}	53.71±15	β	c
I^{130}	51.79±11	β	c
I^{131}	52.23±42	β	c

[a] Ogle, Brown, and Carson, Phys. Rev. **71**, 378 (1947).
[b] B. B. Kinsey and G. A. Bartholomew, Phys. Rev. **89**, 375 (1953).
[c] Hollander, Perlman, and Seaborg, Revs. Modern Phys. **25**, 469 (1953).
[d] C. C. Trail and C. H. Johnson, Phys. Rev. **91**, 474 (1953).
[e] Sher, Halpern, and Mann, Phys. Rev. **84**, 387 (1951).
[f] B. B. Kinsey and G. A. Bartholomew, Can. J. Phys. (to be published).
[g] Bleuler, Stebbins, and Tendam, Phys. Rev. **90**, 460 (1953).
[h] Hanson, Duffield, Knight, Diven, and Palevsky, Phys. Rev. **76**, 578 (1949).

TABLE VI. Constants for isotopic parabolas found by a least squares fit of the formula $\Delta M=\Delta M_0+c(A-A_0)^2$.

Z	Element	A	ΔM_0(mMU)	A_0(amu)	c	No. of points
28	Ni	even	53.22	62.45	−0.342	4
		odd	51.74	62.62	−0.342	4
30	Zn	even	53.27	67.7	−0.21	4
31	Ga	odd	52.54	70.59	−0.14	4
		even	51.64	70.52	−0.27	4
32	Ge	even	55.47	72.87	−0.11	3
		odd	53.33	73.50	−0.11	3
33	As	odd	54.67	76.66	−0.13	3
		even	53.13	76.27	−0.13	3
34	Se	even	58.03	79.46	−0.141	5
		odd	56.38	79.70	−0.141	5
35	Br	odd	58.39	83.28	−0.115	5
		even	56.75	82.82	−0.115	4
36	Kr	even	61.79	85.25	−0.133	5
		odd	60.19	85.61	−0.133	3
46	Pd	even	63.36	104.69	−0.110	5
47	Ag	odd	60.98	107.44	−0.090	4
48	Cd	even	61.58	109.84	−0.094	6
		odd	60.36	109.96	−0.094	5
49	In	odd	59.58	112.62	−0.103	4
		even	58.32	113.26	−0.103	4
50	Sn	even	60.65	116.20	−0.091	7
		odd	59.77	115.78	−0.091	7
52	Te	even	58.9	118.3	−0.055	5
		odd	56.9	119.7	−0.055	4
53	I	odd	55.36	123.70	−0.059	4
		even	53.88	126.05	−0.134	4
54	Xe	even	55.47	127.72	−0.0873	7

We have examined the isotopic sections of the mass surface in detail for elements with Z from 28 to 36 and from 46 to 54. We supplemented masses from this and previous papers[3,4] with masses for unstable atoms listed in Table V. Unstable atomic masses were computed from stable neighbors with the aid of reaction and disintegration data. For 16 elements there is sufficient data to show that parabolas are adequate to fit isotopic mass defect curves. Each element requires two parabolas, one for odd A and one for even, which were determined by the method of least squares. For most elements it is possible to fit similar parabolas, that is parabolas which have the same second-order term. This is a least squares fit of two parabolas with five arbitrary constants which we call a double fit. For all parabolas we demand at least one more point than arbitrary constants; at least 4 points for a single fit and at least 3 points each for a double fit.

The result of this numerical exercise is given in Table VI, which lists the constants for the parabolas expressed as

$$\Delta M=\Delta M_0+c(A-A_0)^2,$$

where ΔM is the mass defect $(A-M)$ in mMU, and ΔM_0 and A_0 are coordinates of the minimum of the isotopic section. Double fits (same c) are indicated by a bracket. Residuals of the fits are usually less than 0.1

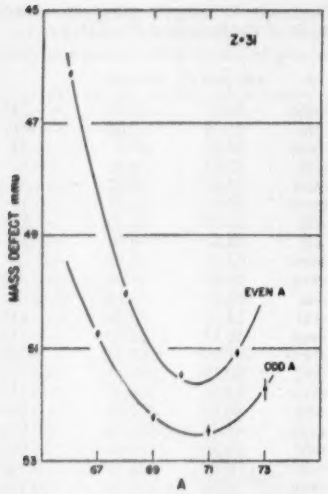

FIG. 2. Parabolas fitted to the mass defects of gallium isotopes.

mMU and are no greater than expected from probable errors. The statistical error of the coefficients in Table VI is difficult to estimate; as a substitute we refer to Figs. 2 to 8 which are representative. (Note the excellent fit of a parabola to the seven, even-A, isotopes of xenon, Fig. 8.) For three elements it was necessary to omit some points. For germanium it was necessary to omit the three isotopes with less than 40 neutrons to obtain a reasonable fit. These three isotopes have greater masses than calculated from the extrapolated parabolas. For indium, an excellent double fit was obtained by omitting In^{115} (Fig. 6). Reaction and disintegration energies do not indicate that the measured mass is incorrect. The fit is never very good for tellurium; however, a double fit is possible if the rare Te^{130} is omitted. Halsted[4] had difficulty measuring tellurium doublets due to the formation of hydrides.

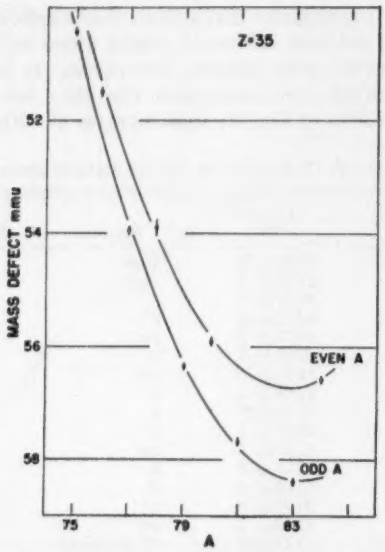

FIG. 4. Similar parabolas fitted to the mass defects of bromine isotopes.

From this examination of isotopic masses we conclude:

1. Parabolas are a good representation of isotopic sections of the mass surfaces within the error of present measurements. Parabolas fit up to and including neutron shells—see Kr^{86} (Fig. 5) and Xe^{136} (Fig. 8).

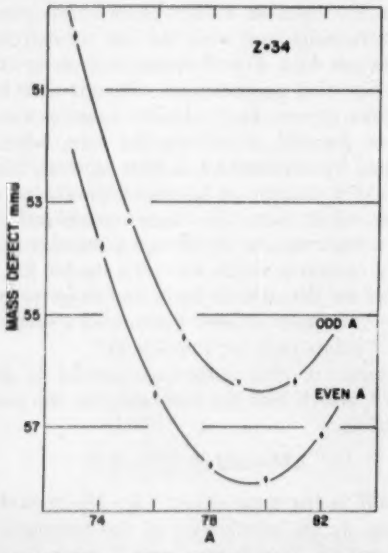

FIG. 3. Similar parabolas fitted to the mass defects of selenium isotopes.

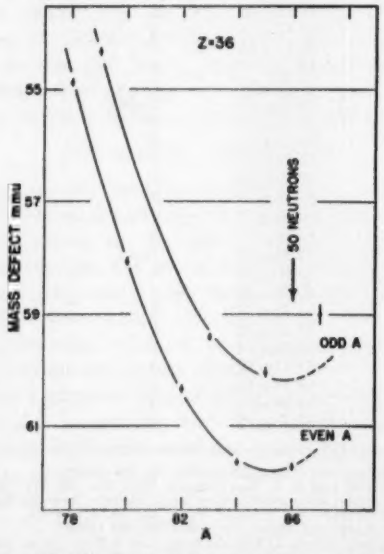

FIG. 5. Similar parabolas fitted to the mass defects of krypton isotopes.

They do not fit across neutron shells—see Kr^{87}_{4}(Fig. 5). We find that Sr^{90} ($52n$) has about 4 Mev less binding energy than predicted by extrapolation of a parabola fitted to three strontium isotopes with 50 or less neutrons.

2. In general, the mass surfaces have the same shape for constant Z, indicated by the success of double fits and by the variation of the constant c. There is no reason to believe that this smooth behavior, between shells, should not also apply to isobars. Exceptions are found immediately above proton shells 28 and 50 (only two examined), where the odd-odd surface has about twice the curvature of the corresponding odd-Z, even-N surface (see Ga, Fig. 2). It is not clear from this form of analysis if 3 surfaces are sufficient.

3. A common assumption is that, for a constant A, the minimum mass occurs at the same Z for each of the mass surfaces. Our analysis shows that this cannot be true for masses discussed in this paper. Unfortunately, direct examination of isobaric parabolas is not useful because there are, at most, two points for each parabola. We reach our conclusion by an examination of minima of isotopic parabolas.

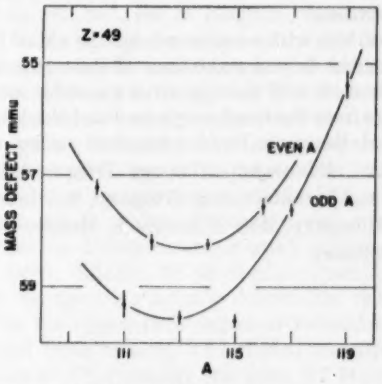

FIG. 6. Similar parabolas fitted to the mass defects of indium isotopes.

The figures show that there is a real difference in position of the minimum A_0 for the two isotopic parabolas for most elements. For example, the ends of the tin parabolas (Fig. 7) show a difference in spacing of about 1 mMU. Table VII lists differences in A_0 and ΔM_0 from Table VI. The difference is given as coordinates of the odd-A surface minus coordinates of the even-A surface. The signs of ΔM_0 differences come from the position of the odd-A surface between the two even-A surfaces.

It is useful to see what these shifts can mean in terms of isobaric parabolas. Consider the "well behaved" region from arsenic to krypton. The constants here vary with sufficient uniformity so that we can apply the following transformations. Let $\partial A_0/\partial Z=b$ and

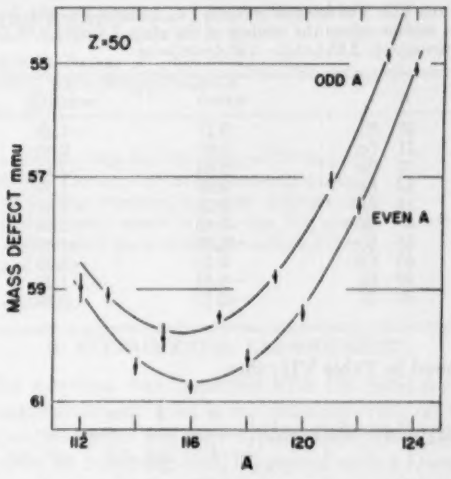

FIG. 7. Similar parabolas fitted to the mass defects of tin isotopes.

$\partial(\Delta M_0)/\partial Z=n$, where A_0, ΔM_0 are the coordinates of the minimum for isotope Z; then the minimum of the constant-A parabola for $A=A_0$ occurs at $Z_0=Z-n/2cb$, with a mass defect $\Delta M_0'=\Delta M_0-n^2/4cb^2$ and a parabolic constant $c'=c/b^2$. As an example we have computed the odd-A parabolas for $A=79$ from data in Table VI. The results are shown in the first two lines of Table VIII. The accuracy is such that we may consider these the same parabola. Line 5 gives a parabola evaluated directly from the three $A=79$ masses (assuming there is one odd-A surface), which shows that the transformation is adequate.

Using the same type of transformation we can calculate, with better accuracy, the difference in the isobaric minima for odd-A and even-A surfaces. Call $a=A_0(\text{odd})-A_0(\text{even})$, and $m=\Delta M_0(\text{odd})-\Delta M_0(\text{even})$,

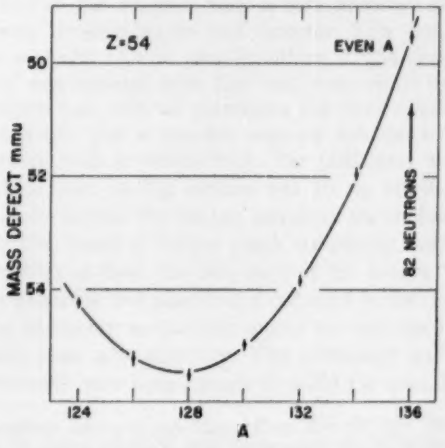

FIG. 8. Parabola fitted to the mass defects of xenon isotopes.

TABLE VII. The isotopic minima (A_0,ΔM_0) of the odd-A mass defect surface minus the minima of the even-A surfaces. A_0(odd) $-A_0$(even)$=a$, ΔM_0(odd)$-\Delta M_0$(even)$=m$.

Z		a(amu)	m(mMU)
28	Ni	0.17	−1.48
31	Ga	0.07	0.90
32	Ge	0.63	−2.14
33	As	0.39	1.54
34	Se	0.24	−1.65
35	Br	0.46	1.64
36	Kr	0.36	−1.60
48	Cd	0.12	−1.22
49	In	−0.64	1.26
50	Sn	−0.52	−0.88

as listed in Table VII; then

$$Z_0(\text{odd})-Z_0(\text{even})=-\frac{a}{b}-\frac{1}{2cb}\frac{\partial m}{\partial Z}+\frac{n}{cb^3}\frac{\partial a}{\partial Z},$$

and

$$\Delta M_0'(\text{odd})-\Delta M_0'(\text{even})=m-(na/b).$$

The spacing m and the shift a for a pair of surfaces does not change fast enough with Z ($\delta Z=2$) to cancel the $-a/b$ term. Lines 3 and 4 of Table VIII list the result of these transformations for our example, $A=79$. Of course, there are no nuclei on an even-A surface at $A=79$, but there is a parabola on a continuous surface. For these isobars the only change necessary in the usual assumptions is to displace the minimum for the odd A parabola 0.1 lower in Z from the even parabolas. This shift is still significant. The transformation has increased the quadratic constant by a factor b^2 ($b\approx 3$) so that the effect on isobaric masses is the same magnitude as on isotopic masses.

TABLE VIII. Isobaric parabolas for $A=79$, computed from isotopic constants for the four surfaces, expressed as $\Delta M=\Delta M_0'+c'(Z-Z_0)^2$.

		Z_0	$\Delta M_0'$(mMU)	c'
even Z	odd N	34.58	56.69	−1.17
odd Z	even N	34.5	56.63	−1.3
even even	—even odd	0.10	1.81	
odd odd	—odd even	0.12	−1.76	
odd A	(3 masses)	34.57	56.55	−1.10

It is apparent that any mass formula that does not include this displacement of the mass surfaces, and current formulas do not, cannot fit the odd and even masses of an isotope within ½ Mev. The change in sign of the displacement for indium and tin suggest strongly that the effect is related to nuclear shell structure.

4. The results generally show a lack of smoothness in the neighborhood of gallium and germanium (see, for example, Table VII). This could be caused by a subshell at 40 neutrons.

The authors wish to acknowledge the aid of Clayton F. Giese, who helped make some of the measurements. The construction of the apparatus was aided materially by grants from the Graduate School and the Minnesota Technical Research Fund subscribed to by General Mills, Inc., *Minneapolis Star and Tribune*, Minnesota Mining and Manufacturing Company, Northern States Power Company, and Minneapolis Honeywell Regulator Company.

Internal Bremsstrahlung from P^{32}†

MAX GOODRICH AND WILBUR B. PAYNE
Physics Department, Louisiana State University, Baton Rouge, Louisiana

(Received January 11, 1954)

The spectral distribution of the internal bremsstrahlung associated with the beta decay of P^{32} has been measured with a NaI scintillation spectrometer from 80 to 900 kev. The shape of the experimental spectrum is compared with the theoretical distribution, corrected for detector response, for three different beta interactions: (1) allowed; (2) first-forbidden; scalar; (3) first-forbidden; tensor; $\Delta J=2$, yes. The results show that of these interactions, only the allowed case gives a theoretical curve within the estimated error (10 percent) of the experimental points over the entire range of energy.

I. INTRODUCTION

THE theory of the energy distribution and intensity of the internal bremsstrahlung (I.B.) accompanying beta decay, given independently by Bloch[1] and Knipp and Uhlenbeck,[2] and extended by Chang and Falkoff[3] to forbidden beta transitions has proved, within the limitations of experiments to date, to be essentially correct. The calculations of Chang and Falkoff as well as those of Bolgiano, Madansky, and Rasetti[4] indicate that the shape of the I.B. spectrum depends only slightly on the degree of forbiddenness of the beta transition. In general this dependence is negligible at the low energies, increases to a maximum at intermediate energies and decreases again as the transition energy is approached, as we show later.

Thus far the measurements on the I.B. spectrum have been made at the low-energy end where the dependence on forbiddenness is small. Consequently, it has been difficult to determine unambiguously whether the spectrum actually follows the theoretical curves for the appropriate degree of forbiddenness as determined from beta-ray and related measurements. In the case of P^{32}, (beta-ray end point 1.7 Mev) which is known to have an allowed shape for the beta spectrum and is designated an *L*-forbidden decay,[5] the I.B. spectrum has been carefully examined from 3 kev to 250 kev by Bolgiano, Madansky, and Rasetti,[4] by Renard[6] and by Novey.[7] Within the accuracy of the experiments, the I.B. spectrum fits the allowed case.

In the present paper the authors report measurements of the I.B. spectrum of P^{32} out to 900 kev and compare the shape of the curve with calculations for the allowed and first-forbidden cases out to these energies where the difference in the two theoretical curves should be experimentally measurable.

† Research supported by: (1) a Frederick Gardner Cottrell grant from The Research Corporation; (2) a grant-in-aid from the Louisiana State University Council on Research.

[1] F. Bloch, Phys. Rev. **50**, 272 (1936).
[2] J. K. Knipp and G. E. Uhlenbeck, Physica **3**, 425 (1936).
[3] C. S. Wang Chang and D. L. Falkoff, Phys. Rev. **76**, 365 (1949).
[4] Bolgiano, Madansky, and Rasetti, Phys. Rev. **89**, 679 (1953).
[5] L. W. Nordheim, Revs. Modern Phys. **23**, 322 (1951).
[6] G. A. Renard, J. phys. radium **14**, 361 (1953).
[7] T. B. Novey, Phys. Rev. **89**, 672 (1953).

II. EXPERIMENTAL ARRANGEMENT

The spectrum was measured with the same scintillation spectrometer used in our previous work on P^{32}.[8] It is of the Jordan and Bell[9] type and employs a 1½ in. diameter by 1 in. long NaI(Tl) crystal with a Dumont K1186 photomultiplier. The pulses, after amplification in a linear amplifier, are analyzed with a single-channel pulse-height analyzer. Gamma rays from Cs^{137} (661 kev) and Ba^{131} (122 kev, 369 kev, and 496 kev) were used to calibrate the linear energy scale.

The high specific activity source used (~0.025 milligram phosphorous per millicurie) was obtained from the Oak Ridge National Laboratory where it was produced by $S^{32}(n,p)P^{32}$, and chemically processed. It was received in the form of phosphate in weak HCl. About 10 mC was placed within a 0.5 cm diameter circle on polystyrene tape of 2.2 mg/cm² thickness and the liquids evaporated. This gave a source thickness of the order of 5 mg/cm². Any trace of continuous radiation from an impurity would have only a small effect on the measured continuum. No evidence of nuclear gamma rays could be observed during the measurements. For these reasons, the source was deemed to be of sufficient purity for the measurements desired. In order to reduce the ratio of external (E.B.) to internal bremsstrahlung, the source was mounted 20 cm above the face of the detector with a beta absorber about half-way between source and detector. This arrangement is similar to that used by others.[4,7] The detector crystal was shielded from light and stray fields by an aluminum can with an aluminum foil cover. Resting on this can was a wooden support holding a lead collimator with a conical hole. The collimator was 5 cm thick, and its top surface was 10 cm above the face of the crystal. On the top surface of the collimator was a ¼-in. sheet of Lucite which completely shielded the collimator from the beta rays of the source. The entire assembly was placed on a tall stool in the center of the laboratory so that the source was not less than 3½ feet from any structure. The collimator and its Lucite cover were large enough to shield the stool from

[8] Goodrich, Levinger, and Payne, Phys. Rev. **91**, 1225 (1953).
[9] W. H. Jordan and P. R. Bell, Nucleonics **5**, No. 4, 30 (1949); W. H. Jordan, Ann. Rev. Nuclear Sci. **1**, 207 (1952).

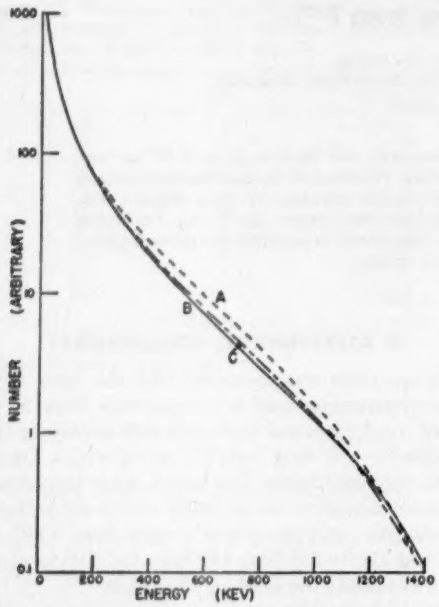

FIG. 1. Calculated energy distributions of internal bremsstrahlung resulting from three beta-interactions (A) allowed; (B) first-forbidden; scalar; (C) first-forbidden; tensor; $\Delta J=2$, yes. The ordinates are normalized to 100 at 150 kev. The calculations are made from equations given in reference 4.

direct rays from the source. The mounting of the source was such that the only betas stopped in a position to see the crystal were those stopped in the Lucite above the collimator hole and a negligible number which moved off at right angles to the collimator axis and were stopped in the polystyrene tape or in the source material. Measurements of the spectrum made with Al foils immediately above the source showed that the contribution due to E.B. produced in the source was negligible.

The contribution of external bremsstrahlung (produced in the beta stopper) to the observed spectrum was determined by obtaining the photon spectrum with an aluminum beta stopper and a copper beta stopper as well as with the Lucite. Extrapolation to $Z=0$ of the number of photons detected as a function of the Z of the beta absorber for various energies, showed that, with the Lucite beta stopper, roughly seven percent of the spectrum was contributed by the external bremsstrahlung. In the calculations this contribution was neglected. This is all the more acceptable since the I.B. and E.B. have a somewhat similar shape.[8]

TABLE I. Calculated relative intensity of internal bremsstrahlung as a function of photon energy for various beta interactions.

k(Mev)	Allowed	First-forbidden scalar	First-forbidden tensor
0.026	7540	7790	7770
0.077	2250	2270	2270
0.179	776	745	749
0.281	397	360	368
0.485	146.7	118.9	126.1
0.690	62.5	46.6	50.9
0.843	32.8	24.1	26.3
1.05	12.33	9.76	10.12
1.20	4.94	4.42	4.27
1.354	1.451	1.51	1.36
1.507	0.591	0.827	0.790

III. DATA AND CALCULATIONS

We show in Table I our calculated I.B. spectrum for: (a) allowed beta transitions, (b) first-forbidden beta transitions; scalar interaction (c) first-forbidden beta transitions; tensor interaction; special case of selection rule ($\Delta J=2$; yes). These calculations are made from the equations given by Bolgiano *et al.*[4] and have been normalized to 1000 at 150 kev. They are also shown graphically in Fig. 1. In the region below 150 kev, the three I.B. spectra are so nearly the same shape that they have been drawn as one line. Above this energy the spectrum for allowed transitions becomes relatively greater than the spectra for the first-forbidden transitions until in the neighborhood of 800 kev, the curves again approach each other and actually cross at about 1400 kev. At 800 kev the separation of the curves reaches almost 30 percent. It will be noted that of the first-forbidden transitions whose I.B. spectra were computed, the tensor interaction spectrum is the more like the spectrum for allowed transitions.

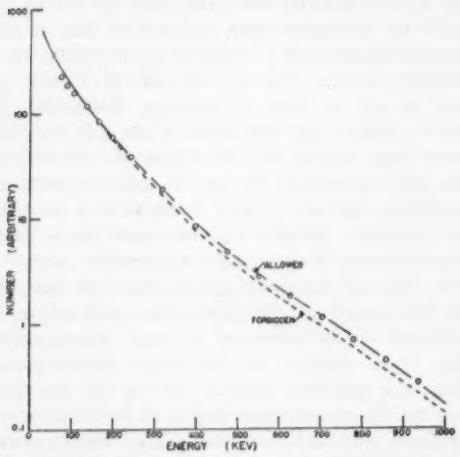

FIG. 2. Energy distribution of internal bremsstrahlung from P^{32}. The dashed curves are the calculated spectra for allowed and first-forbidden (tensor) beta interactions shown in Fig. 1 corrected for detector response. The circles are the experimental points. Standard error from counting statistics is less than the size of the points. Calculated curves and experimental data have been normalized to 100 at 150 kev.

In Fig. 2 we compare the experimental data with the theory. Here the circles are the experimental data, the long dashed curve and the short dashed curve are the theory for the allowed and first-forbidden tensor

interaction cases, respectively, all normalized to 100 to 150 kev. The effect of the detector response has been applied to the theory in this figure. This was done as discussed in our earlier paper[5] except that in the present case, degradation of the photons caused by their passage through the low-Z Lucite absorber was neglected. The uncertainty in the detector response is about 10 percent and forms the principle uncertainty. The standard error from counting statistics is about 3 percent.

The fit of the experimental data to the theory curve for the allowed transition case is much better than 10 percent. We have chosen to show the tensor interaction case for the spectrum from first-forbidden transitions since of the two calculated, it lies nearer the spectrum for allowed transitions. It is, *a fortiori*, evident that the P^{32} I.B. spectrum does not agree with the first-forbidden scalar.

No measurements were made on the ratio of I.B. photons to beta particles. Other experimenters[4,6,7] have found satisfactory agreement working in the region below 250 kev.

We can improve our earlier figures on the upper limit of the possible number of nuclear gammas per beta particle by a factor of 5. If nuclear gammas were present to the extent of 10^{-4} at 200 kev, 4×10^{-5} at 500 kev, or 2×10^{-6} at 900 kev per beta particle, they would have been observed as humps in the spectrum above the smooth curve actually found.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the interest and guidance of J. S. Levinger concerning the theory; the work of Col. R. B. Emerson in preparing the samples; and the use of P. R. Bell's data on the crystal response.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Radiative Electron Capture in Fe^{55}†

L. MADANSKY AND F. RASETTI
The Johns Hopkins University, Baltimore, Maryland
(Received January 14, 1954)

The internal bremsstrahlung emitted in electron capture by Fe^{55} was studied from 14 kev to the upper energy limit at 220 kev. Fair agreement with the formula of Morrison and Schiff is observed at high energies, whereas a strong rise of the intensity occurs at low energies. The discrepancy may be due to failure of the theory to take into account the effects of nuclear charge, capture of p electrons, and forbiddenness of the transition.

THE internal bremsstrahlung emitted in electron capture by Fe^{55} was studied by Maeder and Preiswerk,[1] and by Bell, Jauch, and Cassidy.[2] The high-energy part of the spectrum was shown to agree at least approximately with the formula calculated by Morrison and Schiff[3] for s-electron capture in allowed transitions, assuming a decay energy slightly above 200 kev. The low-energy portion was not investigated in detail owing either to poor resolution of the spectrometer or to presence of impurities.

We therefore undertook careful measurements of the entire spectrum. The Fe^{55} source was over two years old, hence all the Fe^{59} originally present had decayed. Sodium iodide crystals 10 to 12 mm thick were used in combination with DuMont 6292 photomultipliers. The resolution attained corresponds to full widths at half-maximum of 34-36 percent for a 22-kev line and 18-20 percent for an 87.5-kev line. Spectra were recorded under various geometries and with different crystals and photomultipliers, yielding consistent results.

To obtain the spectral distribution of the gamma radiation from the recorded pulse-size spectrum, one must correct for: (1) Percentage of pulses in photoelectric peak; however, up to 200 kev this correction is small at the crystal thicknesses employed. (2) Incomplete efficiency of the crystal at the higher energies. (3) Absorption by materials between source and crystal. (4) Escape of the K radiation of iodine. (5) Gaussian spread of the pulses. Furthermore, at very low energies, one must be certain that none of the pulses observed are due to the K x-rays of Mn at 5.9 kev, at least 10^4 times stronger than the integrated continuum. The curve due to the x-rays, measured by using a thin Fe^{55} source in immediate contact with an NaI crystal, showed that under our experimental conditions the effect of the x-rays was certainly negligible above 14 kev. Hence the bremsstrahlung was investigated only above this energy. Corrections for absorption at low energies in 0.188 g/cm² Lucite and 0.015 g/cm² Al interposed between source and detector were made by means of tabulated absorption coefficients and checked by taking spectra with additional absorbers. Corrections for lack of efficiency at high

† Work supported by the U. S. Atomic Energy Commission.
[1] D. Maeder and P. Preiswerk, Phys. Rev. 84, 595 (1951).
[2] Bell, Jauch, and Cassidy, Science 115, 12 (1952).
[3] P. Morrison and L. I. Schiff, Phys. Rev. 58, 24 (1940).

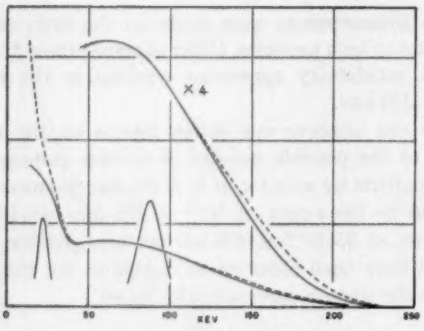

FIG. 1. Experimental spectrum of Fe^{55}. Solid line represents observed pulse-size spectrum; dotted lines indicate spectrum corrected for absorption at low energies and incomplete crystal efficiency at high energies. Peaks obtained from 22-kev and 87.5-kev lines under same conditions show resolution.

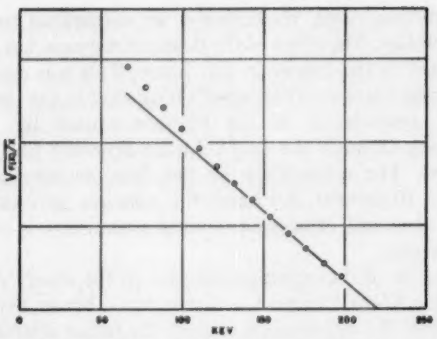

FIG. 2. Plot of $F(x)/x^{\frac{1}{2}}$. Circles represent corrected experimental points; straight line corresponds to formula of Morrison and Schiff with $k_0=220$ kev.

energies were made on the same basis. Corrections for effects (1) and (4) were not attempted. Correction for the Gaussian pulse distribution was applied near the upper energy limit, where it is important for evaluating the decay energy. For the remainder of the energy region above 50 kev, this correction is unimportant; at lower energies it was not attempted owing to lack of a satisfactory theoretical formula and the inaccuracy of the experimental data due to absorption effects. Pulses were detected up to 235 kev; however, taking into account the Gaussian spread of the pulse size, it appears that the maximum energy k_0 is near 220 kev. The latter value is used in comparison with the theory.

The experimental results are represented in Fig. 1. The solid line gives the measured spectrum, the dotted line indicates the corrections for absorption and efficiency. It is immediately apparent that there is a large low-energy component not predicted by the formula of Morrison and Schiff, $f(x)=x(1-x)^2$, where $x=k/k_0$. At the higher energies (above 60 kev) the experimental spectrum corresponds fairly closely, but not exactly, to this formula. To facilitate comparison, in Fig. 2 we plot the square root of the corrected experimental intensity divided by $x^{\frac{1}{2}}$. If the spectrum followed the Morrison-Schiff formula, the points would lie on a straight line. Discrepancies are apparent below 130 kev, but the points above this energy fit well a straight line with $k_0=220$ kev.

As possible effects contributing to the disagreement between the observed spectrum and the Morrison-Schiff formula, the following may be listed.[4] (1) The spectrum was calculated for allowed transitions; the spectra emitted in forbidden transitions are expected to depend on the form of beta interaction. The transition in Fe^{55} is probably first forbidden, since $\log(ft)=6.1$. (2) The contribution of radiative capture of p electrons may be important. Cutkosky[4] evaluated the effect for allowed transitions, neglecting the effect of nuclear charge, and obtained spectra rising at low energies, similar to the one observed. (3) Important corrections at low energies might result from the use of correct Coulomb wave functions even in the case of s electrons; all existing calculations were based on a plane-wave approximation.

In conclusion, it seems impossible at present to compare the experimental results, especially at low energies, with a satisfactory theoretical formula.

[4] We are indebted to Dr. R. E. Cutkosky for communicating to us unpublished results of his calculations and for valuable discussions.

Neutrons from the Proton Bombardment of N^{14}†

FAY AJZENBERG, *Department of Physics, Boston University, Boston, Massachusetts and Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, Massachusetts*

AND

WOLFGANG FRANZEN,* *Department of Physics, Princeton University, Princeton, New Jersey*

(Received January 14, 1954)

The $N^{14}(p,n)O^{14}$ reaction has been investigated at 30, 60, and 150 degrees in the laboratory system by the method of proton recoils in thick nuclear emulsions. The energy of the incident protons was 17.3±0.1 Mev, and the thickness of the melamine target was 0.2 Mev. The results indicate a ground-state Q value of −6.03±0.2 Mev, yielding a mass defect for O^{14} of 12.2 Mev, in good agreement with beta-decay results. The data also indicate levels in O^{14} at excitation energies greater than 5.5 Mev.

I. INTRODUCTION

IN the past few years, extensive studies have been made of the level structures of the isobars[1] with $|T_z|=\frac{1}{2}$, where $T_z\equiv(N-Z)/2$. Wherever sufficient evidence is available, a one-to-one correspondence of the levels in $T_z=+\frac{1}{2}$ and $T_z=-\frac{1}{2}$ mirror nuclei is observed, when nonnuclear effects, such as the n-p mass difference and the difference in electrostatic energies, are taken into account.[2] This correspondence is believed to be a good indication of the charge symmetric character of nuclear forces.

Presumably, corresponding states should also occur in pairs of isobars which differ in the substitution of two protons for two neutrons, that is, $T_z=+1$ and $T_z=-1$ nuclei. No levels have been observed in nuclei with $T_z=-1$ until the present data were obtained, and thus this isobaric comparison has not heretofore been possible. Usually the most direct way of studying such proton-abundant nuclides in the region of the light nuclei is by means of (p,n) reactions which frequently have rather negative Q values. At this time, thick photographic emulsions (in particular, Ilford C-2 plates) are the only means of measuring fairly accurately neutron energies in the Mev range. Unfortunately, while it is possible by the method of proton recoils in thick plates to measure neutron energies to 0.5 percent (using a thin target and a well-defined incident beam), the group resolution is poor. The minimum interlevel distance for group resolution depends on the neutron energy as well as on the target thickness and beam energy spread and, in the best cases, is approximately 7 percent of the neutron energy. It was clear that the Princeton cyclotron external beam width (0.15 Mev) and the thick target (0.2 Mev), necessary for intensity reasons, would result in poor group resolution, but we felt that any insight gained in the behavior of $T_z=-1$ nuclei might nevertheless be of interest.

† This work has been supported in part by the United States Air Force through the Office of Scientific Research and in part by the joint program of the U. S. Office of Naval Research and the U. S. Atomic Energy Commission.

* Now at the University of Rochester, Rochester, New York.

[1] F. Ajzenberg and T. Lauritsen, Revs. Modern Phys. **24**, 321 (1952).

[2] T. Lauritsen, Ann. Rev. Nuclear Sci. **1**, 67 (1952).

O^{14} was studied in order to check the value of the mass of O^{14} derived from beta-decay studies,[3–5] and because some estimate of the magnitude and direction of the energy shift for corresponding levels in the $T_z=+1$ (C^{14}) and $T_z=-1$ (O^{14}) nuclei might be possible. There has been some indication[6] for a level at approximately 4 Mev in C^{14}. Assuming charge symmetry, a level in O^{14} should then appear near 4 Mev. Similarly, the more reliable levels in C^{14} (6.10,[6,7] 6.78,[8] and 6.98 Mev[8]) should have analogs in O^{14}.

II. EXPERIMENTAL PROCEDURE AND RESULTS

The experimental setup is indicated in Fig. 1. The external proton beam from the Princeton cyclotron bombarded a Melmac target placed at the center of a 12-in. diameter evacuated scattering chamber. The energy of the incident protons was determined, prior to the actual exposure, by measuring their range in aluminum. This range was found to be $(442.3\pm2)\times10^{-4}$ g/cm² of aluminum, which is equivalent[9–11] to a proton energy of 17.3±0.1 Mev.

Dr. K. G. Standing prepared the Melmac foil. The particular type of Melmac used (No. 1079 made by the American Cyanamid Corporation) is believed to be 50 percent melamine ($N_6C_3H_6$) and 50 percent filler (probably cellulose). For future experiments, Melmac 404, which is pure melamine, would be preferable.[12] The thickness of the foil was measured to be 2×10^{-3} in.,

[3] R. Sherr and J. B. Gerhart, Phys. Rev. **91**, 909 (1953).

[4] Sherr, Muether, and White, Phys. Rev. **75**, 282 (1949).

[5] J. R. Penning and F. H. Schmidt, Bull. Am. Phys. Soc. **29**, No. 1, 42 (1954).

[6] R. G. Thomas and T. Lauritsen, Phys. Rev. **88**, 969 (1952). However, Bent, Bonner, and Sipple, Phys. Rev. **91**, 472 (1953), do not confirm this result.

[7] Sperduto, Holland, Van Patter, and Buechner, Phys. Rev. **80**, 769 (1950).

[8] A. Sperduto (private communication).

[9] J. H. Smith, Phys. Rev. **71**, 32 (1947).

[10] E. L. Hubbard and K. R. MacKenzie, Phys. Rev. **85**, 197 (1952).

[11] H. Bichsel and R. F. Mozley, Phys. Rev. **90**, 354 (1953) and a very helpful private communication from Dr. R. F. Mozley.

[12] We are very much indebted to Professor Standing for preparing the target and for his comments, which are partially reproduced in the above paragraph.

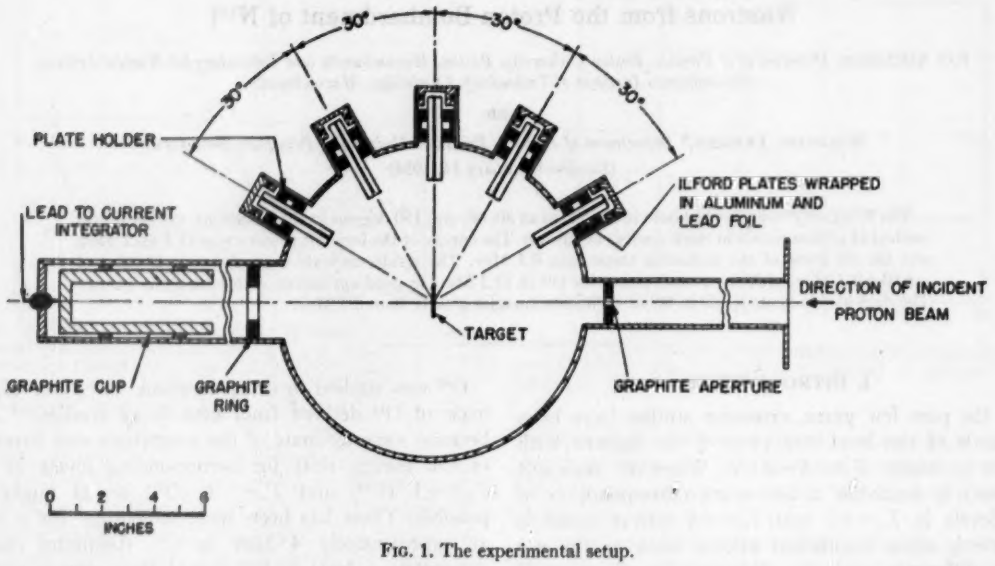

FIG. 1. The experimental setup.

which corresponds to an energy loss of approximately 0.2 Mev for 17-Mev protons.[13]

Ilford C-2 plates, 200 and 400 microns thick, were mounted at five different angles to the incident beam in the evacuated chamber. The plates were wrapped in 2-mil thick aluminum foil for protection from light, and lead sheets 35 mils thick shielded the plates from scattered protons and other charged particles.

The proton beam was collimated by graphite apertures and was stopped in the graphite cup indicated in Fig. 1. As a result, background radiation reached the emulsions from the following sources: Gamma rays from $C^{12}+p$ and from $C^{13}+p$ (presumably mostly from the former since C^{13} has an isotopic abundance of 1.1 percent) and neutrons from $C^{13}+p$. Graphite was chosen because the $C^{13}(p,n)N^{13}$ reaction has a threshold above the bombarding energy used in this experiment (20.0 ± 0.1 Mev).[14]

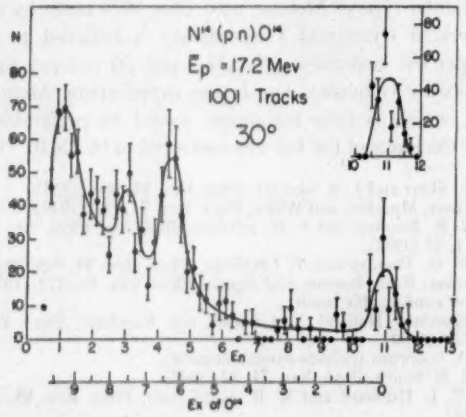

FIG. 2. 30° data. N is the relative number of neutrons per 200-kev interval. $\bar{E}_p$ is the average proton energy.

The gamma rays fogged the plates to such an extent that the maximum exposure possible was 600 microcoulombs, which turned out to be somewhat less than optimum for efficient scanning. While the C^{13}/C^{12} ratio was small, the total amount of graphite seen by the beam was appreciable, and it was feared that the $C^{13}(p,n)N^{13}$ reaction ($Q=-3.003$ mev)[15] would contribute a continuous background of neutrons. This background turned out to be very small (see Figs. 2 and 3). This is reasonable in view of the small solid angle subtended by the photographic emulsions at the graphite cup and apertures, and on account of the angular acceptance criteria for the proton recoil tracks which favored the acceptance of recoils caused by neutrons from the target.

The Melmac target also contained a small amount of C^{13}; the ratio of C^{13} to N^{14} nuclei was approximately 1/100. This might contribute several approximately monoenergetic neutron groups corresponding to the ground state and the first few excited states of N^{13}, if the $C^{13}(p,n)N^{13}$ cross section were fairly high. The upper limit to the number of proton recoil tracks at 30, 60, and 150 degrees which could be ascribed to the ground-state transition is 7, while a total of 2892 tracks were measured. The term "upper limit" is used because the ground-state neutron group from the reaction $N^{15}(p,n)O^{15}$ ($Q=-3.487$ Mev[1]) may be confused with the ground-state neutron group from $C^{13}(p,n)N^{13}$ ($N^{15}/N^{14}=1/250$).

[13] Aron, Hoffman, and Williams, U. S. Atomic Energy Commission Report AECU-663, 1949 (unpublished).

[14] L. Alvarez, Phys. Rev. 75, 1815 (1949).

[15] Richards, Smith, and Browne, Phys. Rev. 80, 524 (1950).

The plates were processed in a manner suggested by Dr. A. H. Armstrong[16] and then scanned with a binocular Leitz microscope[17] using a 100× oil-immersion objective and 10× eyepieces. The criteria for the measurement of the recoil-proton tracks have been discussed previously.[18] The range-energy relation used was that derived by Rotblat.[19] The data shown on Figs. 2 and 3 have been corrected for geometry[20] and for variation of the n-p scattering cross section with neutron energy.[21]

Plates exposed at 30, 60, and 150 degrees to the incident beam were scanned for proton-recoil tracks. Figures 2 and 3 indicate the data at the two smaller angles. The insets in these figures show the appearance of the ground-state groups when additional ground-state tracks were measured. The insets at each of the angles represent fifty ground-state tracks. The ground-state Q value, derived from the results at the three angles, is -6.03 ± 0.2 Mev. The ground-state neutron group width is greater than that calculated on the basis of target thickness, beam width, straggling, the size of the source, and the scanning area and multiple scattering effects. The group width is also larger than predicted from an extrapolation of thin target, "monoergic" beam results.[22] However, it is not felt that the low points at $E_n=11.1$ Mev (30°) and 10.3 Mev (60°) are significant. The resultant high-energy groups, if the points were to be believed, would have a width somewhat smaller than that warranted by the resolution of this experiment.

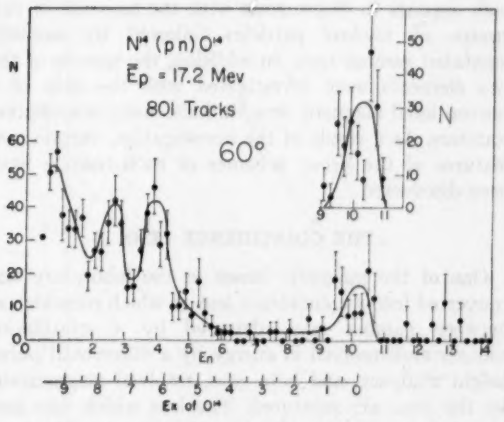

FIG. 3. 60° data. N is the relative number of neutrons per 200-kev interval.

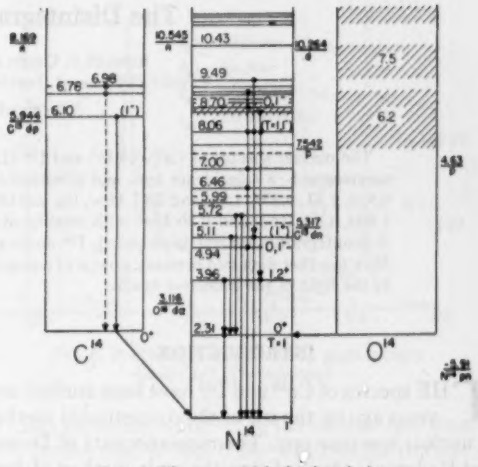

FIG. 4. The mass-14 isobaric triad.

The region between the ground state and an excitation of approximately 5.5 Mev seems to be free of levels unless the corresponding neutron groups are low in intensity at the various angles (less than ¼ of the intensity of the ground-state neutron group). The data for $E_n\lesssim5$ Mev indicate energy levels in O^{14}. It is thought that the peaks corresponding to excitation energies of 6.2, 7.5, and 9.3 Mev are due either to very wide levels, or to a number of unresolved levels in O^{14}. Figure 4 is an attempt to show how the information derived from this work fits into the known picture of the mass-14 triad.[1,3,23] The dashed areas in O^{14} show the upper and lower limits of the regions in which wide or unresolved levels may be located.

In conclusion, we are in agreement with the mass defect for O^{14} predicted from beta-decay results.[1,3–5] Our value for the mass defect is 12.2 ± 0.2 Mev. We do not find a state in O^{14} analogous to the 4-Mev state[6] in C^{14}, unless it has been shifted upwards in O^{14} by about 1.5 Mev. If the group at an excitation energy of 6.2 Mev corresponds to the 6-Mev levels in C^{14} and not to the 4-Mev state, then the energy shift of the corresponding levels in the mirror nuclei O^{14} and C^{14} appears to be small.

We are very much indebted to J. G. Likely and F. C. Shoemaker for their assistance in the exposure of the plates, and to K. G. Standing who prepared the melamine target. The kind interest and support of P. T. Demos and W. W. Buechner of M.I.T. and of M. G. White of Princeton rendered possible this joint experiment.

[16] A. H. Armstrong (private communication) and Allred, Armstrong, and Rosen, Phys. Rev. **91**, 90 (1953).
[17] The microscope was kindly loaned by the Department of Physics, Smith College, Northampton, Massachusetts.
[18] Johnson, Laubenstein, and Richards, Phys. Rev. **77**, 413 (1950).
[19] J. Rotblat, Nature **167**, 550 (1951).
[20] H. T. Richards, Phys. Rev. **59**, 796 (1941).
[21] R. K. Adair, Revs. Modern Phys. **22**, 249 (1950).
[22] See, for instance, F. Ajzenberg, Phys. Rev. **82**, 43 (1951).
[23] Private communications from D. H. Wilkinson, A. Sperduto, and R. J. Mackin.

The Disintegration of Co^{55} and I^{130}†

ROBERT S. CAIRD AND ALLAN C. G. MITCHELL
Physics Department, Indiana University, Bloomington, Indiana
(Received December 2, 1953)

The nuclear spectra of Co^{55} (18 hr) and I^{130} (12.6 hr) have been studied with the help of a magnetic lens spectrograph, a coincidence lens, and scintillation counters. Co^{55} emits gamma rays having energies 0.477, 0.935, 1.41, 0.253, 1.84, and 2.17 Mev, the last three being quite weak; and four beta-ray groups of energies 1.500, 1.03, 0.53, and 0.26 Mev with relative abundances of 53.3, 39.5, 4.9, and 2.3 percent, respectively. A disintegration scheme is proposed. I^{130} emits gamma rays of energy 0.409, 0.528, 0.660, 0.744, and 1.15 Mev together with two beta-ray groups of energies 1.02 and 0.597 Mev. The energy-level scheme is discussed in the light of the collective model.

INTRODUCTION

THE spectra of Co^{55} and I^{130} have been studied some years ago by the use of the conventional methods of nuclear spectroscopy. The measurements of Deutsch and Hedgran[1] indicated that the main method of decay of Co^{55} occurs by the emission of two positron groups of energies 1.50 and 1.01 Mev together with the emission of three gamma rays of energies 0.47, 0.935, and 1.41 Mev. They also reported a weak gamma ray of energy 0.095 Mev which does not fit into the scheme.

The spectrum of I^{130} was studied by Roberts, Elliott, Downing, Peacock, and Deutsch[2] in the early days of nuclear spectroscopy. They showed that I^{130} decays by the emission of two beta-ray groups of energies 1.03 and 0.61 Mev together with four gamma rays whose energies are 0.417, 0.537, 0.667, 0.714 Mev. According to these authors the position of the gamma ray at 0.417 Mev is established from the energies of the two beta-ray groups but the remaining three gamma rays are all in cascade and their arrangement is arbitrary. Smith, Mitchell, and Caird[3] measured the spectrum of Cs^{130}, which also goes to Xe^{130}, with the hope of establishing the order of the gamma rays in cascade, but found that these gamma rays were not excited in the disintegration of Cs^{130}.

Since a coincidence lens has recently been put into operation in this laboratory, it was felt that Co^{55} and I^{130} would be suitable subjects for investigation since each appears to disintegrate with the emission of two groups of nuclear particles followed by suitably separated gamma rays. In addition, the spectra of the two elements were investigated with the help of a conventional magnetic lens, and also using scintillation counters. As a result of the investigation, certain new features of the decay schemes of each nuclide have been discovered.

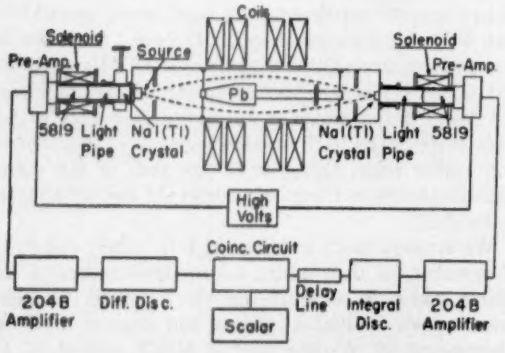

FIG. 1. Schematic diagram of coincidence lens.

THE COINCIDENCE LENS

One of the magnetic lenses in the laboratory was converted into a coincidence lens in which coincidences between gamma rays, detected by a scintillation counter and resolved in energy by a differential pulse-height analyzer, and beta rays, resolved magnetically by the lens, are measured. The lens which was used was that described by Bunker, Canada, and Mitchell.[4] A schematic of the arrangement of counters and electronic components is shown in Fig. 1.

The end plate of the instrument nearest the source was refitted to permit the insertion of a canned NaI(Tl) crystal and a lucite "light-piper" immediately behind the source. The NaI(Tl) crystal was in the form of a cylinder 1 in. in diameter by 1 in. high. The "light-piper" was 4½ in. long. Optical contact between the crystal, "light-piper," and a 5819 photomultiplier tube was made with the help of silicone grease. In order to minimize the effect of the magnetic field of the lens on the photomultiplier tube a solenoid was constructed, which was operated in parallel with the main coils of the lens. Thus, for any current in the main coils, the field at the photomultiplier was essentially zero.

Beta rays were detected at the other end of the lens with the help of a NaI(Tl) crystal in the form of a cube 1 cm on each edge. A "light-piper," 5819 photomultiplier tube, and solenoid, similar to those already described, were employed. The crystal was mounted *in vacuo* by clamping it to a Lucite window which was

† Supported by the joint program of the U. S. Office of Naval Research and the U. S. Atomic Energy Commission.
[1] M. Deutsch and A. Hedgran, Phys. Rev. **75**, 1443 (1949).
[2] Roberts, Elliott, Downing, Peacock, and Deutsch, Phys. Rev. **64**, 268 (1943).
[3] Smith, Mitchell, and Caird, Phys. Rev. **87**, 454 (1952).
[4] Bunker, Canada, and Mitchell, Phys. Rev. **79**, 610 (1950).

fitted with a rubber gasket to hold the vacuum. Optical contact was made between the crystal, Lucite "light-piper," and photomultiplier tube.

The general layout of the electronic equipment is is shown in Fig. 1. The output of each photomultiplier goes through a cathode-follower preamplifier and from there into an Atomic Instrument Company Model 204B linear amplifier. In the beta-ray channel, the integral discriminator on the linear amplifier is used to cut down the noise and shape the pulses. The output then goes through a delay line to the coincidence circuit. In the gamma-ray channel, the output of the linear amplifier is fed to an Atomic Instrument Company Model 510 single-channel pulse-height analyzer. That portion of the pulse spectrum which is selected by the analyzer goes on to the coincidence circuit. The coincidence analyzer consists of suitable pulse-shaping

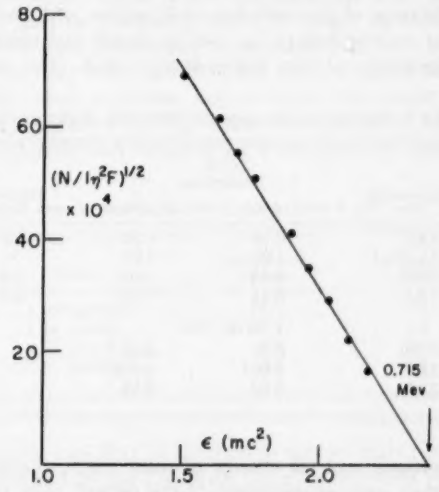

FIG. 2. Interior beta-ray group from Rb^{86} observed with coincidence lens.

devices, a diode discriminator, and amplifying stage, and an integral discriminator. The coincidence circuit was usually run at 0.20-microsecond resolving time. Either singles channel or the coincidence channel could be selected by a three-position switch and the output fed into a fast (1-microsecond dead time) decade scalar. The coincidence system was lined up by using a double pulse generator, putting simultaneous pulses into each of the cathode followers, and adjusting the delay line until true coincidences were obtained.

The singles system was calibrated by measuring the beta-ray spectrum and the internal conversion line from Cs^{137} with the lens and the beta counter. For the adjustments used, the resolution was about 6 percent. The gamma-ray counter was likewise calibrated using the Cs^{137} source. The full width at half-maximum for this line was about 15 percent.

The coincidence apparatus was tested by measuring

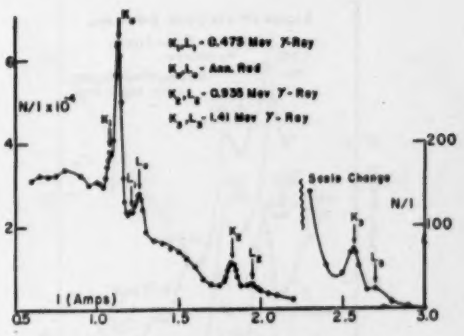

FIG. 3. Photoelectrons ejected by gamma rays from Co^{55} (Pb radiator).

the main spectrum of Au^{198} and the interior spectrum of Rb^{86}. In the case of Au^{198}, the scintillation counter at the gamma ray end was adjusted to accept the 0.411-Mev line, the beta-ray distribution in coincidence with this gamma ray was measured, and a Fermi plot in good agreement with the known decay scheme was obtained. In the case of Rb^{86}, the beta-ray distribution was measured using a single counter to obtain the shape and end point of the high-energy group. In addition, with the gamma-ray counter focused on the line at 1.08 Mev, the coincidence distribution was examined.

The Fermi plot for the interior group is shown in Fig. 2. Assuming an allowed shape for this group, the end-point energy is 0.715±0.010 Mev. The high-energy group exhibits the well-known shape characterized by $\Delta j = \pm 2$ and a change of parity, and the end point being determined as 1.798 Mev. These results are in agreement with the work of Muether and Ridgeway[5] and the earlier work of Zaffarano, Kern, and Mitchell.[6]

Experiments on Co^{55}

Co^{55} (18 hr) was prepared by bombarding iron with 11.5-Mev deuterons on the probe of the Indiana University cyclotron. With the exception of the 9-hour isomeric state of Co^{58}, which gives only low-energy internal conversion electrons, the only other cobalt isotopes produced are long-lived and suitable corrections can be made for them. A chemical separation was made on all source material. The purified cobalt was then examined for beta rays and gamma rays.

THE GAMMA-RAY SPECTRUM

The gamma-ray spectrum was measured in a magnetic lens spectrometer using both lead and uranium radiators. A typical spectrum, taken with a lead radiator of 12-mg/cm² surface density is shown in Fig. 3, in which the number of counts per momentum interval (N/I), is plotted as a function of the current in the coils of the lens. All lines except one at 0.5 amp (not shown) decay

[5] H. R. Muether and S. L. Ridgeway, Phys. Rev. 80, 750 (1950).
[6] Zaffarano, Kern, and Mitchell, Phys. Rev. 74, 682 (1948).

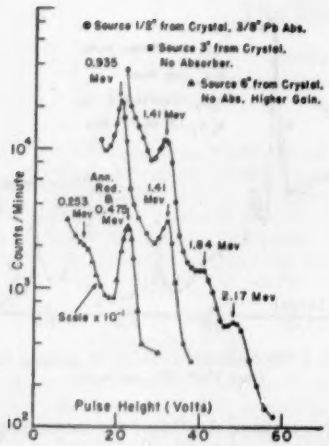

FIG. 4. Scintillation spectrum of gamma rays from Co^{58}.

with a period of 18 hours and are attributed to Co^{58}. The line at 0.5 amp (an L line for a gamma ray of 0.119 Mev), decays with a much longer period and is attributed[7,8] to Co^{57}. The line at 0.475 Mev is not well resolved from annihilation radiation. In order to resolve these two lines and to get an estimate of the relative intensity of the line at 0.475 Mev with respect to the others, a lead radiator of 7.75-mg/cm² surface density was employed and the region around 0.500 Mev was studied carefully. The energies and relative intensities of the gamma rays, and the abundances per positron of the lines, are given in Table I. The results for the strong lines are in reasonable agreement with those of Deutsch and Hedgran.[1]

As a result of the measurement of the energy distribution of the positrons (see below), two weak prositron groups were found in addition to the two main groups reported by Deutsch and Hedgran. These two groups had lower energies than the two main groups. In order to look for low-intensity gamma rays connected with the emission of these positron groups, the gamma rays

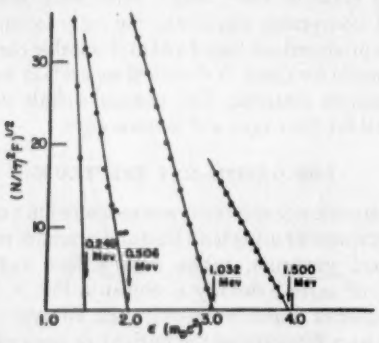

FIG. 5. Fermi plot of positron groups from Co^{58}.

[7] Cheng, Dick, and Kurbatov, Phys. Rev. 88, 887 (1952).
[8] L. G. Elliott and M. Deutsch, Phys. Rev. 64, 321 (1943).

were investigated with the help of a scintillation spectrometer. The scintillation spectrometer consisted of a 1-in. cylindrical NaI(Tl) crystal together with a Dumont 6292 photomultiplier tube. After appropriate amplifications the pulses were analyzed using a single channel pulse height analyzer. The instrument was calibrated with the help of the Cs^{137} and Co^{60} gamma rays.

In searching for the higher-energy gamma rays, a ¾-inch lead absorber was used to absorb the lower-energy gamma rays. This prevented "pulse pileup" from contributing to the general background in the high-energy region. The low-energy region, of course, was measured without lead absorber. The results are shown in Fig. 4. The strong lines found in the magnetic spectrometer were found in the scintillation spectrometer together with three additional lines at 2.17, 1.84, and 0.253 Mev. Owing to the weakness of the lines and the decrease of photoelectric cross section with energy, strong sources had to be used to locate and measure the intensity of the higher-energy lines. The usual

TABLE I. Summary of energies and intensities of gamma-rays.

Line energy Mev	Co^{58} Strong lines I/I_0	I/positron	Deutsch and Hedgran
0.477	0.14	0.28	0.3
AnnRad	1.00	1.00	...
0.935	0.78	1.56	1.4
1.41	0.13	0.26	0.3
	Weak lines		
0.253	0.01	0.02	
1.84	0.003	0.006	
2.17	0.02	0.04	

precautions such as change of solid angle, change of absorber, and measurement of the period, were taken to make sure that the peaks were genuine and not due to pulse addition.

Owing to the large difference in intensity between the strong lines and the weak lines, a series of comparisons had to be made using various sources and different experimental arrangements. For example, the lines at 1.41, 1.84, and 2.17 Mev were compared using a strong source and the lead absorber, corrections being made for the absorber. The line at 0.935 Mev was then compared with that at 1.41 Mev using no absorber. Finally, the lines at 0.935, 0.510 and 0.477, and 0.253 were compared using a weaker source. The results of the measurements are shown in Table I. It is felt that the relative intensities of the weak lines give an order of magnitude only.

THE PARTICLE SPECTRUM

The particle spectrum—positrons and internal conversion electrons—was investigated with the help of three different magnetic spectrometers—a magnetic

lens, the coincidence lens, and a 180° spectrometer. Preliminary experiments gave indications of positron groups of lower energy than the two main groups reported by Deutsch and Hedgran. A careful investigation was made of the particle spectrum in the magnetic lens using a counter window which would transmit particles down to less than 10 kev. The source, made by evaporating active cobalt chloride onto a thin aluminum backing, was 0.5 mg/cm² thick. Aside from the Co^{55} spectrum, two internal conversion lines of long period having energies of 120 and 134 kev, K and L lines for a gamma ray of 23-kev energy, and K lines for gamma rays of 72- and 95-kev energy having half-lives of approximately 9 hours were seen. The lines from the gamma rays at 120 and 134 kev presumably arise from the 270-day[9] Co^{57} and those from the gamma ray at 23 kev from the isomeric transition[9] of Co^{58}. The lines at 72 and 95 kev, having a period of 9 hours, may also be connected with the decay of Co^{58}, but further work is necessary to establish this.

The data were corrected for small amounts of any other cobalt activities, and a Fermi plot made. The results are shown in Fig. 5 and the end points, relative abundances, and values of log ft are shown in Table II.

TABLE II. Positron groups from Co^{55}.

E(Mev)	Abundance percent	log ft
1.500±0.008	53.3	6.3
1.03 ±0.03	39.5	5.8
0.53 ±0.03	4.9	5.9
0.26 ±0.02	2.3	6.0

It will be seen that, in addition to the two main positron groups at 1.50 and 1.03 Mev, there are two weak groups at 0.53 and 0.26 Mev. A similar experiment made with a 180° spectrometer, to investigate the internal conversion lines of Co^{55}, gave results in agreement with those given above. The values given in Table II are the averages of all the experiments.

An experiment was then made with the coincidence lens. The scintillation spectrometer was set on the high-energy side of the peak of the 1.41-Mev gamma ray and those positrons which were in coincidence were measured. The end point obtained from a Fermi plot of this spectrum, shown in Fig. 6, is 1.03 Mev. The weak low-energy groups are not shown.

Finally, the internal conversion of the strong lines at 0.477, 0.935, and 1.41 Mev were investigated in a 180° spectrograph. The intensity of the internal conversion lines were found to be 1.10×10^{-4}, 1.24×10^{-4}, and 3.53×10^{-5} internal conversion electrons per positron for the lines in order of increasing energy.

[9] Nuclear Data, National Bureau of Standards, Circular No. 499 (U. S. Government Printing Office, Washington, D. C., 1950).

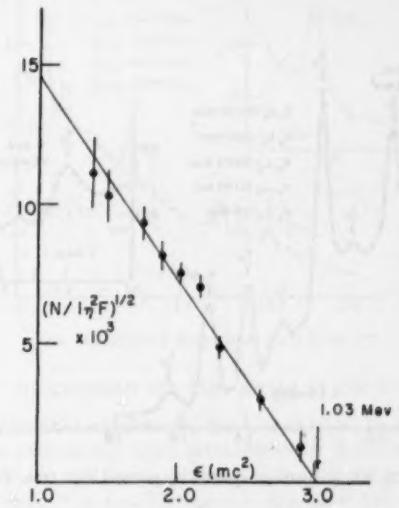

FIG. 6. Fermi plot of second group of positrons in Co^{55} taken with coincidence lens.

CONCLUSIONS

The results of the investigation are embodied in the decay scheme shown in Fig. 7. The strong gamma rays at 0.477, 0.935, and 1.41 Mev, together with the strong positron groups at 1.50 and 1.03 Mev, constituting the principal mode of decay, are in agreement with the work of Deutsch and Hedgran.[1]

The present experiments show two new positron groups at 0.53 and 0.26 Mev together with additional weak gamma rays. If the strong gamma rays and the positron groups are used to established the energies of the levels, the weak gamma rays agree both in energy

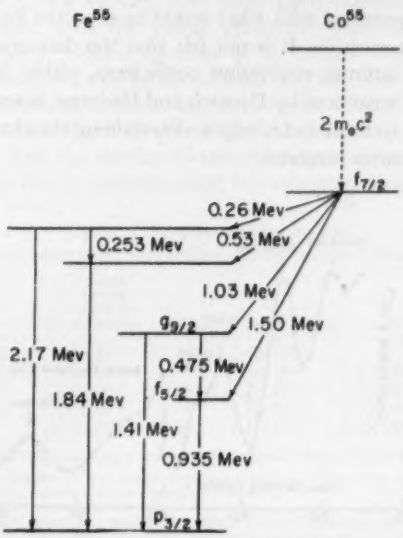

FIG. 7. Disintegration scheme for Co^{55}.

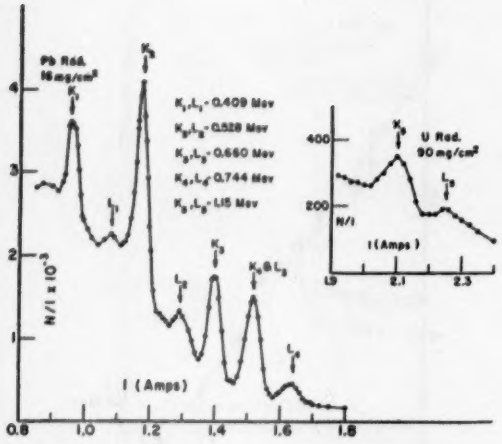

FIG. 8. Photoelectrons ejected by gamma rays from I^{130}.

and intensity with the transitions as shown in the level scheme. Two other possible gamma rays involving transitions from the state at 1.90 Mev to those at 1.41 and 0.938 Mev were not seen. The first of these transitions would be obscured by the annihilation radiation and the second by the line at 0.938 Mev.

According to the shell model, the single particle level for the ground state of $_{27}Co^{55}$ is $f_{7/2}$ and for the ground state of $_{26}Fe^{55}$ is $p_{3/2}$. Thus it appears reasonable that no positron transition takes place to the ground state of the product. The level order in Fe^{55} should be $p_{3/2}$, $f_{5/2}$, $g_{9/2}$. The experimentally determined value of log ft for the 1.50-Mev positron group is 6.3 and appears to be somewhat high for a $f_{7/2}$ to $f_{5/2}$ transition. The value of log ft for the 1.03-Mev group is 5.8 and is in good agreement with what would be expected for a $f_{7/2}$ to $g_{9/2}$ transition. It is not felt that the determination of the internal conversion coefficients, either by the present writers or by Deutsch and Hedgran, is accurate enough to be of much help in determining the character of the states involved.

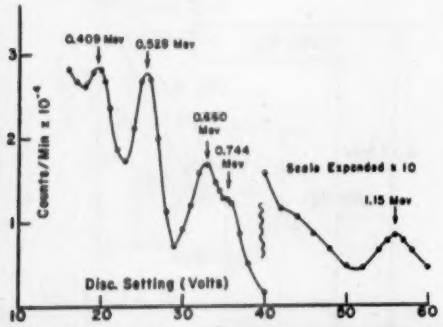

FIG. 9. Scintillation spectrum of gamma rays from I^{130}.

Experiments on I^{130}

I^{130} was prepared by bombarding tellurium with deuterons. The sample was dissolved in HNO_3 with some KI, to serve as a carrier, and the iodine was distilled off, extracted by CCl_4 and eventually precipitated as AgI for use as sources. The I^{130} (half-life 12.6 hrs) was accompanied by other iodine activities and corrections for these had to be made in doing the experiments. In order to keep unwanted activities to a minimum, separated Te^{130} was used as a target for some experiments.

THE GAMMA RAYS

In order to investigate the gamma rays, the spectrum of the photoelectrons ejected from either a lead or a uranium radiator was measured in a magnetic lens spectrograph. The main curve of Fig. 8 shows the number of photoelectrons per unit momentum interval for a lead radiator of 16 mg/cm² surface density plotted against the current in the lens. The peaks labelled K_1, L_1, to K_4, L_4, denote the K and L photoelectric peaks arising from four different gamma rays emitted by I^{130}. The peak labelled K_4+L_3 contains the K photoelectrons for a line at 0.744 Mev and the L peak for a line at 0.660 Mev. A similar experiment with a uranium radiator resolved the K_4 from the L_3 peak but caused the overlapping of other lines. The energies of these lines are given in Table III.

The gamma rays were also measured using a NaI(Tl) scintillation spectrometer and the results are shown in Fig. 9. In this figure, the two low-energy lines shown in Fig. 9 are clearly resolved but the line at 0.660 Mev shows some overlapping with that at 0.744 Mev. In addition a higher-energy line of energy about 1.15 Mev is seen. Experiments were performed to be sure that this line was not due to "pulse pileup." The spectrum was then reinvestigated in the magnetic lens spectrometer employing a strong source and a uranium radiator. The result is shown in the insert of Fig. 8 where a K and an L photopeak will be seen corresponding to a line at 1.15 Mev.

Estimates of the relative intensities of the gamma rays were obtained by determining the area under the several K photopeaks involved and applying the corrections for the variation of photoelectric efficiency with energy. Since the K photopeak for the line at 0.744 and the L photopeak for the line at 0.660 coincide, a correction had to be made for the intensity of the

TABLE III. Gamma rays from I^{130}.

Gamma ray	Energy of gamma ray Mev	Relative intensity
1	0.409	0.3
2	0.528	1.00
3	0.660	0.9
4	0.744	0.8
5	1.15	0.4

latter. This was accomplished by measuring the K and L photopeaks for the line at 0.530 Mev and using this to compute the contribution of the L peak for the line at 0.660 Mev. The intensity of the line at 1.15 Mev was determined by comparing the K photopeak produced by it in a uranium radiator with those of the other lines produced by the same source using a uranium radiator. The results, shown in Table III, are not felt to be known to better than 20 percent.

THE BETA-RAY SPECTRUM

The distribution of beta rays and internal conversion electrons was investigated with the help of the magnetic lens spectrometer. Figure 10 shows that part of the spectrum containing the internal conversion lines. All lines are internally converted. A Fermi analysis of the beta-ray spectrum was made and it was found to be resolvable into two groups with end-point energies of 1.02 and 0.597 Mev, in agreement with Roberts *et al.* No groups of higher energy than that at 1.02 Mev were found. Table IV gives the characteristics of the beta-ray groups.

The number of internal conversion electrons per disintegration were obtained by taking the area under the K internal conversion peaks and comparing this to the area under the beta-ray spectrum. For the line at 1.15 Mev the L line was not resolved from the K line. The results are given in Table V together with estimates of the K/L ratios.

COINCIDENCE EXPERIMENTS

Two scintillation counters using NaI(Tl) crystals and equipped with differential pulse-height analyzers were used to measure coincidences between certain gamma-ray lines. The resolving time of the coincidence circuit was 0.2 microsecond.

Two sets of experiments were carried out. In the first one, Channel 2 was set on the high side of the peak arising from the (0.660+0.744)-Mev lines, and Channel 1 was swept through the range of the spectrum. Figure 11 shows the result of the experiment. It will be seen that the coincidence curve shows a peak at 0.660 Mev, somewhat to the lower energy side of the composite peak shown in the singles spectrum. Peaks are also seen for the 0.528- and 0.409-Mev lines in the coincidence spectrum.

A similar experiment was carried out in which Channel 1 was set on the line at 1.15 Mev and Channel 2 swept across the spectrum. The results show that the 1.15-Mev line is in coincidence with those at 0.660 and 0.528 Mev but not with those at 0.409 and 0.744 Mev.

TABLE IV. Beta rays from I^{130}.

E Mev	Rel. abundance Percent	log ft
1.02	46	6.5
0.597	54	5.5

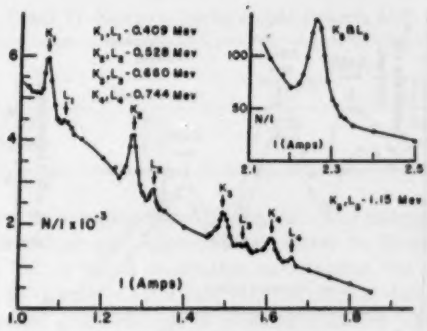

FIG. 10. Internal conversion lines from I^{130}.

DISCUSSION OF THE RESULTS ON I^{130}

The present experiment has shown that, in addition to the gamma-ray lines established by Roberts *et al.*, there is an additional gamma ray having an energy of 1.15 Mev. The position of the line at 0.409 Mev is fixed by the difference in energy of the two beta-ray groups and is therefore placed at the top of the disintegration scheme in agreement with previous work. The existence of the line at 1.15 and the gamma-gamma coincidence experiments involving this line show that it must be in parallel with those at 0.409 and 0.744 Mev, which definitely fixes the position of the 0.744-Mev line. The other two lines are placed in cascade at the bottom of the scheme. Relative intensity determinations, although not very accurate, tend to confirm this assignment. The proposed decay scheme is shown in Fig. 12. The line at 0.528 Mev is placed below that at 0.660 Mev from considerations involving the collective model.

In order to obtain some idea of the spins and parities of the states involved, the internal conversion coefficients of the various lines have been calculated and compared with the values given by Rose, Goertzel, and Perry.[10] Since the lines at 0.528 and 0.660 Mev are in cascade and follow all beta and all other gamma transitions, the calculation is straightforward. In order to calculate the internal conversion coefficients for the other lines, the number of transitions per disintegration for the line in question must be calculated making use

TABLE V. Internal-conversion data.

E_γ	K internal-conversion electrons per disintegration	K/L	α_K(obs)	α_K(theor)
0.409	4.4×10^{-4}	11	1.6×10^{-4}	...
0.528	5.5×10^{-3}	8	5.5×10^{-3}	$6.8\times10^{-3}(E2)$
0.660	3.2×10^{-3}	16	3.2×10^{-3}	$3.8\times10^{-3}(E2)$
0.744	2.1×10^{-3}	4	2.7×10^{-3}	$2.9\times10^{-3}(E2)$
1.15	$(8.8\times10^{-5})^a$	..	2.5×10^{-4}	$4.7\times10^{-4}(E1)$

[a] The number in parenthesis denotes that this of the total number of internal conversion electrons per disintegration.

[10] Rose, Goertzel, and Perry, Oak Ridge National Laboratory Report ONRL-1023, June 25, 1951, (unpublished).

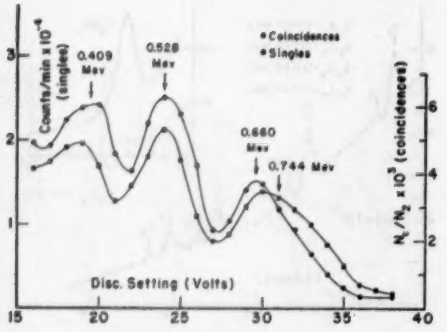

FIG. 11. Gamma-gamma coincidence experiment. (Channel 2 Set on the line at 0.744 Mev.)

of the assumed disintegration scheme, the relative abundance of the two-beta-ray groups, and the branching ratio for the gamma rays emitted from the state at 2.34 Mev. The branching ratio,

$$X = I_{0.407}/(I_{0.407}+I_{1.15}),$$

can be calculated either from the observed relative intensities of the lines at 0.407 and 1.15 Mev or from the relative intensities of the 0.407- and 0.744-Mev lines taking into consideration the relative abundance of the two beta-ray groups. The value for the branching ratio is $X=0.47\pm0.04$. The number of transitions per disintegration involving the line at 0.744 Mev is therefore

$$N^{0.744}/N_{dis}=0.54X+0.46=0.79.$$

Taking into account the number of K internal-conversion electrons per disintegration given in Table V, the internal conversion coefficient can be calculated. A similar procedure was used for the other lines. The results together with the theoretical values are shown in Table V. The internal conversion coefficients for lines at 0.528, 0.660, and 0.744 Mev all agree with the theoretical prediction for $E2$ radiation. The internal conversion coefficient of the 0.409- and 1.15-Mev lines are much less certain since their determination is very sensitive to the branching ratio X. The line at 1.15 Mev appears to be reasonably consistent with $E1$ radiation, while that at 0.409 Mev might be $E1(4.2\times10^{-3})$, $E2(4.0\times10^{-2})$, or $M1(1.74\times10^{-2})$. Since the ground state of Xe^{130} has zero spin and even parity, the $E2$ character of the lines at 0.528, 0.660, and 0.744 Mev, suggests spins of 2, 4, 6, and even parity for the first three excited states. This is in agreement with what is to be expected from either the single-particle model for an even-even nucleus or from the collective model. The value of log ft for the 1.02-Mev beta group would imply $\Delta I=0$, yes for this transition.[11] Combining this information with that obtained for the first three excited states of Xe^{130}, the spin and parity of the ground state of I^{130} is 6, odd. The value of log ft for the 0.597-Mev group shows that it is allowed and is consistent with $\Delta I=1$, no. The spin and parity of the state in Xe^{130} at 2.34 Mev would then have to be 5, odd, which is consistent with the $E1$ character of the 1.15-Mev line. The internal conversion coefficient and relative intensity of the line at 0.409 Mev is still somewhat of an anomaly.

Since neither the parent nor the product nucleus corresponds to a case in which the proton or the neutron number is one less or one greater than a closed shell, the single-particle model is not strictly applicable. It is interesting to compare the present results with what would be expected on the basis of the collective model of Bohr and Mottelson.[12,13] They point out that in the region between closed shells, the strong-coupling collective model gives good results for the ratio of the energies of the lower excited states. The closer one approaches closed shells (magic numbers for protons or neutrons) the less appropriate the strong-coupling theory becomes. In the present instance both the parent and product are reasonably close to closed shells (50 protons and 82 neutrons). Nevertheless, the present experiments may serve as a test of the collective model.

For even-even nuclei with sufficiently great deformation, the collective model predicts a rotational spectrum of levels with energies:[12]

$$E_I=(\hbar^2/2J)I(I+1)+\Delta E_I;\quad I=0,2,4,\cdots, \qquad (1)$$

in which J is the effective moment of inertia, and ΔE_I, the first order correction to the strong-coupling formula, is given approximate y by

$$\Delta E_I=-2(\hbar\omega)^{-2}(\hbar^2/J)^3I^3(I+1)^2. \qquad (2)$$

Here $\hbar\omega$ is the surface phonon excitation energy, about 2.2 Mev at $A=130$. Equation (2) is the sum of Eq.

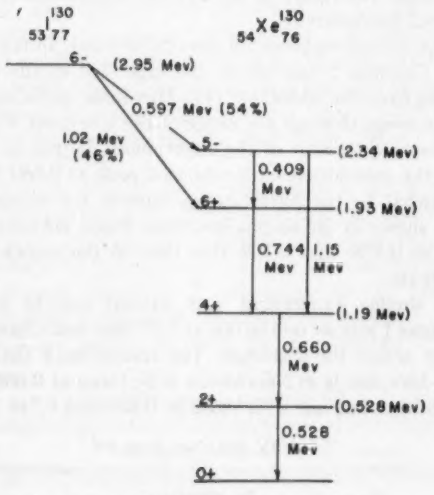

FIG. 12. Disintegration scheme of I^{130}.

[11] L. W. Nordheim, Revs. Modern Phys. **23**, 322 (1950).

[12] A. Bohr and B. R. Mottelson, Kgl. Danske Videnskab. Selskab., Mat.-fys. Medd. **27**, No. 16 (1953).

[13] A. Bohr and B. R. Mottelson, Phys. Rev. **90**, 717 (1953).

(VI 3a) and (VI 3b) of reference 12, in which ω_β and ω_γ have been replaced by limiting strong-coupling values.

One may check the rotational interpretation of the low states of Xe^{130} qualitatively in several ways: (1) the level order is consistent with the predicted order 0+, 2+, 4+, 6+. (2) One can calculate the nuclear deformation from the energy of the first excited state using the relation[14]

$$\beta^2 = 171/A^{5/3}E_2(\text{Mev}). \quad (3)$$

The result is $\beta = 0.310$, which fits very well into the systematic trend of the other Xe isotopes,[14] suggesting a regularly decreasing nuclear deformation as the closed neutron shell at $N = 82$ is approached. (3) The ratios of the energies of the excited states, E_4/E_2 and E_6/E_2 deviate from the limiting strong coupling ratios in the expected direction, but the deviation is too large to receive quantitative explanation in terms of the first order correction given by Eq. (2). The comparison of theoretical and experimental ratios is shown in Table VI. It seems reasonable to interpret the Xe^{130} levels as collective rotations, but the nuclear deformation is not great enough to lend quantitative validity to the strong coupling approximation which works so well in the rare earth region and the region near uranium.

TABLE VI. Energy ratios for excited states in Xe^{130}.

	Experiment	Theory: Without ΔE_I	Theory: With ΔE_I
E_4/E_2	2.25	3.33	1.94
E_6/E_2	3.66	7.00	...

The authors are indebted to Dr. M. B. Sampson and the cyclotron group for making the bombardments, and to Mr. Arthur Lessor for making the chemical separations. They are also indebted to Dr. Kenneth W. Ford for many helpful discussions. The separated Te^{130} was obtained from the Oak Ridge National Laboratory.

[14] K. W. Ford, Phys. Rev. **90**, 29 (1953).

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

Angular Correlations in the $Be^9(d,p\gamma)Be^{10}$ and $Be^9(d,\alpha\gamma)Li^7$ Reactions*

L. COHEN,† S. S. HANNA, AND C. M. CLASS‡
Department of Physics, The Johns Hopkins University, Baltimore, Maryland
(Received October 29, 1953)

The $(d,p\gamma)$ correlation at $E_d = 0.84$ Mev was measured under the following conditions: (1) proton direction fixed at 0° to the deuteron beam, gamma direction varying, (2) proton direction fixed at 90° to the beam, gamma direction varying in the deuteron-proton plane, and (3) proton direction fixed at 90°, gamma direction varying in the 90° plane, perpendicular to the beam. The three correlations were different, and showed varying amounts of anisotropy. The appearance of terms in $\cos^4\theta$ in these correlations indicates that the spin of the first excited state of Be^{10} is ≥ 2. Coupled with other information on this state, this result gives a spin assignment of 2. A qualitative discussion of the correlations is given, and the effect of the deuteron stripping process is considered. The $(d,\alpha\gamma)$ correlation was also measured with the alpha direction fixed at 90° to the beam and the gamma direction varying in the deuteron-alpha plane. Failure to detect a departure from isotropy again confirms the assignment of $\frac{1}{2}$ to the spin of the first excited state of Li^7.

I. INTRODUCTION

THE electromagnetic radiation from the first excited state of Be^{10} has been studied in coincidence with short-range protons in the reaction, $Be^9(d,p)Be^{10*}(\gamma)Be^{10}$. The coincidence yield in such a reaction is, in general, a function of considerable complexity, involving the directions of the deuteron, proton, and gamma ray, as well as the deuteron energy and the properties of the various states and particles taking part in the process. The correlation function will be designated by $W(\theta_p,\theta_\gamma,\phi)$, where θ_p and θ_γ are measured with respect to the direction of the deuteron, and ϕ is the dihedral angle between the proton-deuteron and gamma-deuteron planes, all in the center-of-mass coordinate system.

The following correlation functions were measured:

(1) $W(0,\theta_\gamma,\phi) = W(0,\theta_\gamma)$. If only waves of the same parity are present in the gamma-ray transition, as will be assumed henceforward, the correlation function has reflection symmetry in the $\theta_\gamma = 90°$ plane, and can be written:

$$W(0,\theta_\gamma) = \sum_{n=0}^{K} A_{2n}\cos^{2n}\theta_\gamma = W(0,180-\theta_\gamma),$$

* Assisted by a contract with the U. S. Atomic Energy Commission.
† Now at Knolls Atomic Power Laboratory, General Electric Company, Schenectady, New York.
‡ Now at Rice Institute, Houston, Texas.

where K is an integer which cannot be greater than I_2 or L_γ.[1,2] I_2 is the spin of Be^{10*} and 2^{L_γ} is the multipolarity of the gamma radiation.

(2) $W(90,\theta_\gamma,180)$. If one assumes compound-nucleus formation, the participating states of the compound nucleus play an important role in determining the form of the correlation function.[3] If only states of the same parity in the compound nucleus take part, the correlation function will be symmetric about $\theta_\gamma=90°$, as before:

$$W(90,\theta_\gamma,180)=\sum_{n=0}^{K} B_{2n}\cos^{2n}\theta_\gamma=W(90,180-\theta_\gamma,180),$$

where K is again limited by I_2 or L_γ. In addition,

$$W(90,\theta_\gamma,0)=W(90,\theta_\gamma,180),$$

since $W(90,\theta_\gamma,0)$ is equivalent to $W(90,-\theta_\gamma,180)$. If, however, there are states of opposite parity, antisymmetric terms are introduced. Designating these terms by W', one obtains,

$$W'(90,\theta_\gamma,180)=-W'(90,\theta_\gamma,0)=W'(90,180-\theta_\gamma,0).$$

(3) $W(90,90,\phi)$. In general the correlation function is given by

$$W(90,90,\phi)=\sum_{n=0}^{K} C_{2n}\cos^{2n}\phi.$$

If the deuteron beam does not provide an orientation in the reaction then of course a unique correlation is obtained independent of the direction of the proton. If the reaction proceeds by means of the stripping mechanism, the correlation function may be written:[4]

$$W_s(\theta_s)=\sum_{n=0}^{N} D_{2n}\cos^{2n}\theta_s,$$

where θ_s is the angle of the gamma ray measured with respect to the propagation direction of the captured neutron in the laboratory. The integer N cannot be greater than I_2 or L_γ or l_n, the orbital angular momentum of the captured neutron.

A measurement was made also of the alpha-gamma correlation in the $Be^9(d,\alpha)Li^{7*}(\gamma)Li^7$ reaction. Since no anisotropy was detected in the $W(90,\theta_\gamma,180)$ correlation, as expected, no further measurements were made for other alpha directions.

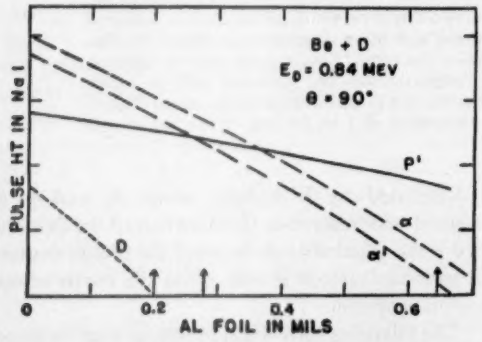

FIG. 1. Approximate pulse height with a NaI(Tl) detector of particles from Be^9+d as a function of absorber thickness.

II. APPARATUS

Sodium iodide crystals were used in conjunction with 5819 photomultiplier tubes for the detection of both the gamma rays and the protons. The crystal used in the gamma ray detector was a one-inch cube; the detector was mounted outside the target chamber and could be rotated about the axis of the chamber. In order to reduce the response of the proton detector to neutrons and gammas, thin sodium iodide crystals were used. A crystal $\frac{1}{16}$ in. thick or less was mounted inside the target chamber on a window optically coupled to an external photomultiplier tube. The crystal was covered by a 0.02-mil nickel foil.

Two cylindrical target chambers were used during the course of the experiment. One chamber was constructed of aluminum with a $\frac{1}{8}$-in. wall; the other of brass with a $\frac{1}{32}$-in. wall. Two sets of targets were employed: one consisted of thin beryllium foils mounted on thin aluminum or nickel backings, the other of thin beryllium deposits on thicker copper or tantalum backings. The thin backings were used in the observations in which the protons were detected in the forward direction. All the targets were approximately 0.07 mg/cm^2.

Pulses from the detectors were amplified and analyzed in single-channel pulse-height analyzers, which were modifications of the circuit designed by Roulston.[5] The outputs of the analyzers actuated blocking oscillators which presented uniform pluses to the coincidence circuit and to the scalers which recorded the single-channel counting rates. The coincidence circuit was a modification of one used previously,[2] with a resolving time of about 10^{-7} sec. The accidental coincidence rate was monitored continuously by means of a second, similar coincidence circuit which received pulses delayed by 0.75 μsec from one of the channels, and underlayed pulses from the other.

[1] Biedenharn, Arfken, and Rose, Phys. Rev. **83**, 586 (1951). Other possible limitations on K are discussed in this and the following reference.

[2] C. M. Class and S. S. Hanna, Phys. Rev. **87**, 247 (1952).

[3] The statements under (2) and (3) follow directly from an examination of the general correlation function obtained from the function given in reference 1 by suitable modification to include multiple states of the compound nucleus.

[4] Biedenharn, Boyer, and Charpie, Phys. Rev. **88**, 517 (1952); L. J. Gallaher and W. B. Cheston, Phys. Rev. **88**, 684 (1952); G. R. Satchler and J. A. Spiers, Proc. Phys. Soc. (London) **A65**, 980 (1952).

[5] K. I. Roulston, Nucleonics **7**, No. 4, 27 (1950).

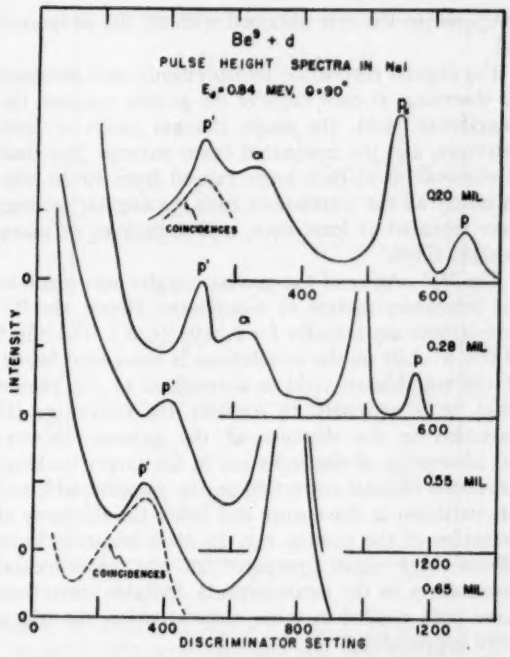

FIG. 2. Pulse height spectra for different absorbers. The two lower curves were taken at increased gain but are plotted to approximately the same energy scale as the two upper curves. The high-energy end of the spectra in the lower curves is badly distorted by over loading in the amplifier. Curves labelled coincidences show the particles in coincidence with 3.4-Mev gamma radiation.

The deuteron beam, obtained from an electrostatic accelerator, had an energy of 0.84 Mev for practically all the measurements.

III. THE $Be^9(d,p)Be^{10*}(\gamma)Be^{10}$ REACTION

The observed charged-particle spectrum included two groups of protons, two groups of alpha particles, and two groups of tritons from Be^9+d, proton groups from $C^{12}+d$ and $O^{16}+d$, and elastic deuterons. In order to avoid overloading the detector and the circuits, it was necessary to eliminate the elastic deuterons by inserting a suitable amount of absorbing foil between target and crystal. Figure 1 gives the approximate pulse heights of the various relevant groups as a function of thickness of aluminum foil. The curves were calculated using the pulse height data on NaI(Tl) of the Illinois group.[6] The target chamber was provided with a foil wheel incorporating four foils, and the spectra obtained are shown in Fig. 2. p, p', and α indicate groups from the beryllium reaction, and p_c is the proton group from carbon. In the two lower curves for the thicker foils, a higher grain was used and the upper end of the spectrum is badly distorted by overloading in the amplifier. In the two upper curves the short-range proton group p' is superimposed on the two alpha groups (not resolved). In the lower curves the alphas are well below the short-range protons. In all the spectra the very broad group of short-range tritons from the beryllium reaction[7] also falls in the region of the p' group.

As in the case of the charged particles, the gamma-ray spectrum is very complex.[8] In addition to the 3.36-Mev gamma from $Be^9(d,p)Be^{10*}$, there are at least seven gamma rays from $Be^9(d,n)B^{10*}$, a 0.87-Mev gamma from $O^{16}(d,p)O^{17*}$, and a 3.1-Mev radiation from $C^{12}(d,p)C^{13*}$. The sodium iodide spectrum of these gamma rays has been investigated in detail in this laboratory, and the results will appear in a separate publication. Under low resolution, one obtains the spectrum shown in Fig. 3. The broad peak contains the multiple structure due to pair production and Compton scattering in the crystal of the following gamma rays: 3.6 and 2.9 Mev from B^{10*}, 3.4 Mev from Be^{10*}, and 3.1 Mev from C^{13*}.

The p-γ coincidences were obtained as follows. The gamma-ray detector was calibrated with radiation of known energy. With a wide, 7.5-volt channel, the gamma-ray analyzer was set to record radiation of about 3.4 Mev. The proton analyzer, also with a 7.5-volt channel, was then set in the region of the short-range proton group, and a coincidence spectrum was obtained by varying the setting of the proton analyzer. Typical coincidence curves obtained are shown in Fig. 2. With the proton channel on the peak of the coincidence curve, a coincidence spectrum was then obtained for the gamma rays. A typical curve is shown in Fig. 3.

Coincidence curves were obtained in this manner before nearly every run. Despite the fact that all these

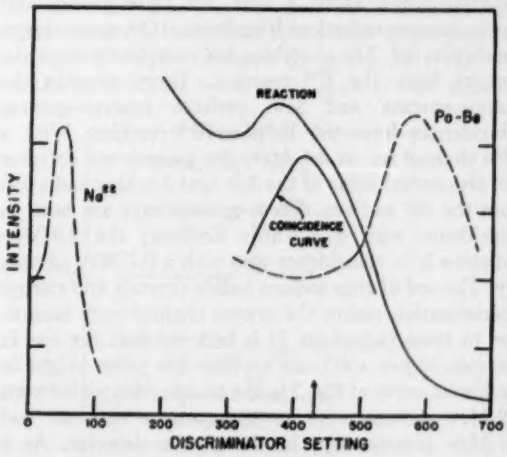

FIG. 3. Pulse height spectra for the gamma radiation from Be^9+d (labelled "reaction"), Na^{22} (0.5 Mev), and Po-Be (4.4 Mev). The coincidence curve shows the radiation in coincidence with short-range protons.

[6] Taylor, Jentschke, Remley, Eby, and Kruger, Phys. Rev. 84, 1034 (1951).

[7] I. Resnick and S. S. Hanna, Phys. Rev. 82, 463 (1951).

[8] Rasmussen, Hornyak, and Lauritsen, Phys. Rev. 76, 581 (1949).

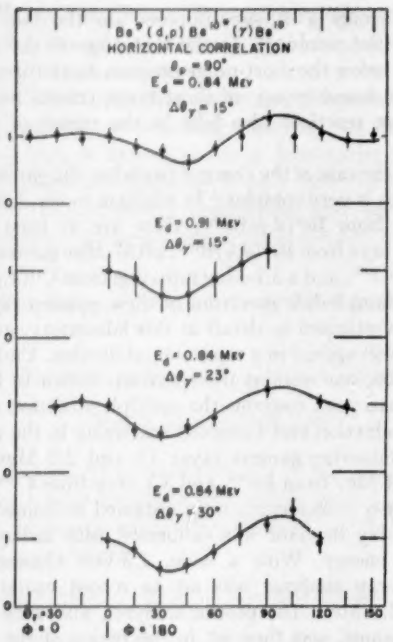

FIG. 4. The $W(90,\theta_\gamma,180)$ and $W(90,\theta_\gamma,0)$ correlations. Measurements were made from $\theta_\gamma=0$ to 45° in the $\phi=0$ plane and from $\theta_\gamma=0$ to 150 in the $\phi=180$ plane. $\Delta\theta_\gamma$ indicates the angular aperture of the gamma counter.

measurements indicated that the proton-gamma coincidences were well isolated, other sources of coincidene radiations may be considered. They are listed in Tablt I. With the channel of the gamma detector set to exclude radiation below about 3 Mev, the alpha-gamma and proton-gamma radiations from Be^9 and O^{16}, respectively, are eliminated. The absorbing foil completely stops the protons from the C^{12} reaction. There remains the gamma-gamma and also perhaps neutron-gamma coincidences from the $Be^9(d,n\gamma)B^{10}$ reaction. With a wide channel set on 3.4 Mev, the gamma-ray detector will also record some of the 2.9- and 3.6-Mev radiation from the B^{10} nucleus. These gamma rays are both in coincidence with a 1.3-Mev neutron; the 2.9-Mev radiation is in coincidence also with a 0.7-Mev gamma ray. The use of thin sodium iodide crystals and energy discrimination makes the proton channel very insensitive to these radiations. It is believed that the rise in the coincidence spectrum at very low pulse height in the lowest curve of Fig. 2 is due to coincidences between 2.9-Mev gamma rays in the gamma detector and 0.7-Mev gamma rays in the proton detector. As a further test of the purity of the desired coincidences, at the maxima of the coincidence spectra, an additional foil was placed between the target and the proton detector so as to completely stop the protons with very little effect on the neutrons and gammas. The coincidence rate obtained in this manner was negligible compared to the rate obtained without the additional foil.

The angular correlation measurements were obtained by observing, at each angle of the gamma counter, the coincidence yield, the single channel yields of both detectors, and the integrated beam current. The time of observation at each angle ranged from 10–30 min. In nearly all the correlations runs the angular settings were repeated at least once, and sometimes as many as eight times.

The 90° setting of the proton counter was made in the laboratory system of coordinates. Hence, the 90° correlations are actually for $\theta_p=95°$(c.m.). The effect of this 5° shift on the correlations is considered below. If the coincidence yield is normalized to the proton yield, it is necessary to consider the following: (1) variation in the distance of the gamma counter; (2) absorption of the radiation in the target backing; (3) center-of-mass correction on the gamma yield, and (4) variation in the energy and hence the efficiency of detection of the gamma ray. In most instances these effects were small compared to the experimental uncertainty in the measurements. Suitable corrections have been applied in those cases in which the effects were appreciable.

(A) The $W(90,\theta_\gamma,180)$ Correlation

The aluminum target chamber was mounted with its axis vertical. The beam, the proton detector (at 90°), and the rotating gamma detector were in a horizontal plane. The target plane was vertical and at an angle with respect to the beam which was varied between 30° and 60° in various parts of the experiment. The aluminum foil thickness was either 0.20 or 0.28 mil producing the charged particle spectra at the top of Fig. 2.

Four measurements of the angular correlation were made under somewhat different conditions. The results are shown in Fig. 4. A pronounced variation of the coincidence yield with angle is observed. The data from the four runs were averaged together and plotted in Fig. 5(A). The correlation appears to have the property that

$$W(90,\theta_\gamma,0)=W(90,180-\theta_\gamma,180).$$

Accordingly, data in the $\theta=0°\rightarrow90°$, $\phi=0°$ quadrant have been plotted at the corresponding points in the

TABLE I. Coincident radiations from deuteron bombardment of Be^9 and target impurities. Particle energies are computed for $E_d=0.84$ Mev, $\theta=90°$ and an absorbing foil of 0.2-mil aluminum

Target	Particle	Gamma ray
Be^9	p', 1.5 Mev	3.4 Mev
Be^9	α', 3.6 Mev	0.5 Mev
Be^9	n, γ	2.9, 3.6 Mev
O^{16}	p', 1.5 Mev	0.9 Mev
C^{12}	p', 0 Mev	3.1 Mev

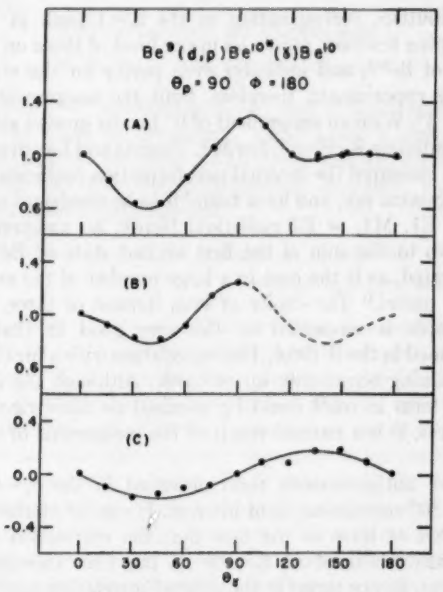

FIG. 5. (A) shows an average of the four runs in Fig. 4. Points in the $\theta_\gamma=0°\rightarrow 90°$, $\phi=0°$ quadrant have been plotted in the $\theta_\gamma=90°\rightarrow 180°$, $\phi=180°$ quadrant. (B) shows the data in (A) folded about $\theta_\gamma=90°$ with corresponding points averaged. (C) shows the result of subtracting the points in (B) from the data in (A).

$\theta=90°\rightarrow 180°$, $\phi=180°$ quadrant, producing a smooth curve in this quadrant. In order to display the symmetric part of this curve, it was folded about $\theta_\gamma=90°$ and corresponding points averaged to produce the symmetric curve in Fig. 5(B). Subtracting the points in this plot from the observed yield, gives the antisymmetric curve in Fig. 5(C).

The data in Fig. 5(B) are fitted very well by the least-square curve,

$$W(90,\theta_\gamma,180)=1.24(1-1.23\cos^2\theta_\gamma+1.03\cos^4\theta_\gamma),$$

which is the curve drawn in the figure. The points in Fig. 5(C) have been fitted with the curve,

$$W'(90,\theta_\gamma,180)=-0.20\sin 2\theta_\gamma,$$

which seems to provide a good description of the data. Since the dependence of the correlation function on the angle ϕ is not known, it is not possible to obtain an exact modification of these functions for the finite apertures of the detectors. An approximate calculation, assuming the dependence on ϕ to be negligible over the apertures, indicates that the coefficients should be be increased by 14–18 percent.

(B) The $W(90,90,\phi)$ Correlation

The brass target chamber was mounted horizontally with its axis along the beam direction. The target was oriented so that its normal made angle of about 60° with the beam and with the direction of the proton counter (at 90° to the beam). The gamma counter was rotated in the vertical $\theta_\gamma=90°$ plane. The aluminum foil was 0.65 mil, producing the spectrum in the lowest curve of Fig. 2.

Two separate runs were made. Their average is shown in Fig. 6. A very small correlation is observed, at most 10 percent. Because this correlation was performed under different conditions, the $W(90,\theta_\gamma,180)$ correlation was remeasured briefly. No great accuracy was achieved, because of the difficulty of making settings and measurements on the gamma counter in the horizontal plane. The results are given in Fig. 6. They are in substantial agreement with the earlier measurements on the $W(90,\theta_\gamma,180)$ correlation.

(C) The $W(0,\theta_\gamma)$ Correlation

The brass target chamber was reoriented with its axis vertical and with the proton detector at 0° to the beam. Beryllium targets on thin backings were used so as to allow protons to reach the detector. With an 0.9 mil absorbing foil, a spectrum similar to the lowest curve of Fig. 2 was obtained. The location of the beam tube and the proton detector limited the settings of the gamma counter to angles between 45° and 150°. In some of the runs the gamma counter was retracted at the extreme angles to obtain greater angular coverage in θ_γ, necessitating a large inverse square correction at these angles.

All the observations made on this correlation are shown in Fig. 7, plotted to the same scale as for the other correlations. A weighted average of all the data is given at the bottom of the figure.

In this correlation there appears to be a small but definite departure from isotropy which is symmetric about $\theta_\gamma=90°$. A satisfactory fit to the data can be obtained by including terms up to $\cos^4\theta$. A least square

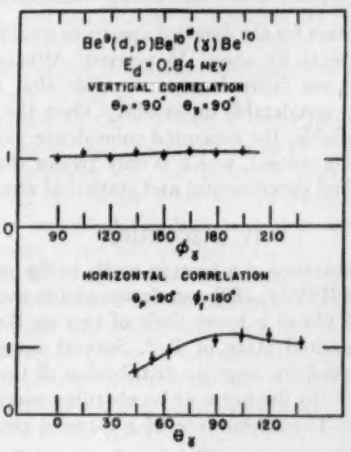

FIG. 6. The $W(90,90,\phi)$ correlation and the $W(90,\theta_\gamma,180)$ correlation taken with the same physical and geometrical arrangement.

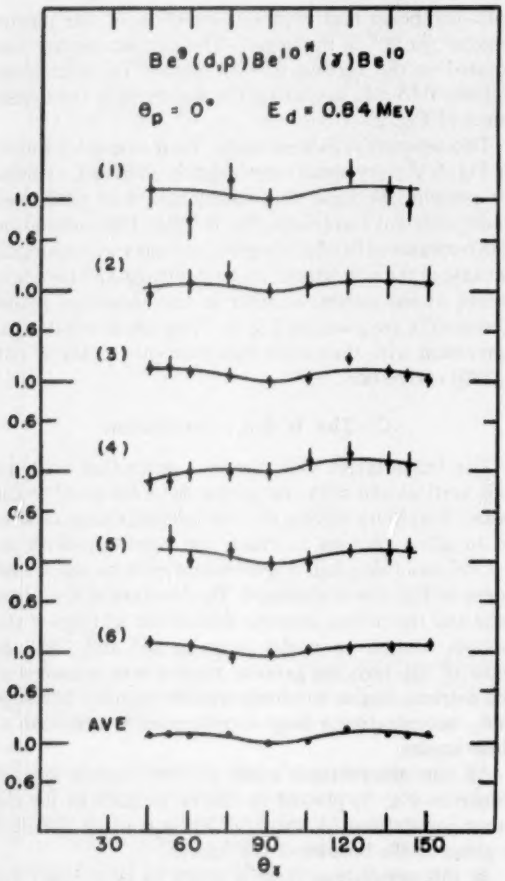

FIG. 7. The $W(0,\theta_\gamma)$ measurements and a weighted average of all the data.

analysis gives

$$W(0,\theta_\gamma)=1.03(1+0.20\cos^2\theta_\gamma-0.23\cos^4\theta_\gamma).$$

The correction for the detector apertures would increase the coefficients by about 15 percent. Although these coefficients are fairly large, it is clear that they are subject to considerable uncertainty. Over the range of angles available, the measured coincidence yield varies by about 10 percent, which is only two or three times the combined experimental and statistical error.

IV. DISCUSSION

The appearance of a term in $\cos^4\theta_\gamma$ in the symmetric part of the $W(90,\theta_\gamma,180)$ correlation and in the $W(0,\theta_\gamma)$ correlation places a lower limit of two on the spin of the first excited state of Be^{10}. Several investigators[9] have observed the angular distribution of the protons with respect to the beam at bombarding energies from 3–15 Mev. The appearance of a forward peak in the distribution, corresponding to the $l_n=1$ peak in the stripping reaction, places an upper limit of three on the spin of Be^{10*}, and indicates even parity for the state. These experiments, therefore, limit the assignment to 2^+ or 3^+. With an assignment of 0^+ for the ground state, the radiation is either E2 or M3. Thomas and Lauritsen[10] have measured the internal pair formation coefficient of this gamma ray, and have found it to be consistent only with E1, M1, or E2 radiation. Hence, an assignment of two to the spin of the first excited state of Be^{10} is indicated, as is the case in a large number of the even-even nuclei.[11] The choice of two, instead of three, for the spin is supported by the very good fit that is obtained in the $W(90,\theta_\gamma,180)$ correlation with a function containing terms only up to $\cos^4\theta_\gamma$. Although the lack of a term in $\cos^6\theta$ could be ascribed to other circumstances, it is a natural result of the assignment of spin two.

The antisymmetric term observed in the $\theta_p=90°$, $\phi=180°$ correlation, is of interest. It can be attributed in part at least to the fact that the correlation was actually obtained at $\theta_p=95°$ in the c.m. coordinate system. Hence terms in the general correlation function containing $\cos\theta_p$ which vanish at $\theta_p=90°$, are small but not negligible at $\theta_p=95°$. These terms arise from interference among magnetic substates in the triple correlation process[1] and provide a dependence on θ_γ, having the symmetry observed, although not necessarily of the simple form $\sin 2\theta_\gamma$.[12] As mentioned in the introduction, the same behavior, even at $\theta_p=90°$, can be ascribed formally to the presence in the compound nucleus of two or more participating states, not all of the same parity. The proton and neutron angular distributions from Be^9+d, at bombarding energies in the range of the present investigation, show asymmetries which can be attributed to interfering states in the compound nucleus.[7,13] These same distributions also give evidence for some deuteron stripping at these bombarding energies. It is therefore interesting to examine the stripping angular correlation in connection with the measurements in the present investigation.

It is clear that the three correlations that were studied are not compatible with the stripping process alone. The stripping correlation should be a function only of the angle between the gamma ray and the direction of recoil of the Be^{10*} nucleus. When the protons are emitted at 0°, the axis of the stripping correlation coincides with the proton direction. When the protons are observed at 90°, the axis turns out to be at 45° to the beam, in the $\phi=180°$ plane.

A stripping angular correlation can be computed for each channel spin entering the stripping process. With

[9] F. A. El Bedewi, Proc. Phys. Soc. (London) **A65**, 64 (1952); C. F. Black, Phys. Rev. **87**, 205 (1952); Fulbright, Bruner, Bromley, and Goldman, Phys. Rev. **88**, 700 (1952).

[10] R. G. Thomas and T. Lauritsen, Phys. Rev. **88**, 969 (1952).

[11] G. Scharff-Goldhaber, Phys. Rev. **90**, 587 (1953).

[12] *Note added in proof.*—We have recently measured the correlation for a setting in the laboratory corresponding to $\theta_p=90°$ in the center-of-mass system. The antisymmetric term appears to be reduced, but not eliminated, at this setting.

[13] Pruitt, Hanna, and Swartz, Phys. Rev. **87**, 534 (1952).

$l_n=1$,

$$W_s(\theta_s)=1+\cos^2\theta_s \text{ for channel spin 1,}$$

$$W_s(\theta_s)=1-\cos^2\theta_s \text{ for channel spin 2.}$$

In the $\theta_p=0°$ correlation, $\theta_s=\theta_\gamma$, and it is not possible, on the basis of symmetry, to separate a stripping correlation of the form $1+D_2\cos^2\theta_s$ from the observed correlation. The stripping process, however, cannot account for a term in $\cos^4\theta$.

In the $\theta_p=90°$, $\phi=180°$ correlation, $\theta_s=\theta_\gamma-45°$, and $\cos^2\theta_s=\frac{1}{2}(1+\sin 2\theta_\gamma)$.

Hence the term in $\sin 2\theta_\gamma$ observed in this correlation could be attributed in part at least to the stripping process. $W_s(\theta_s)$ would have the form $1+D_2\cos^2\theta_s$, with $D_2<0$. Removal of a correlation of this type from the $W(90,90,\phi)$ and $W(0,\theta_\gamma)$ correlations would in both cases increase the observed departure from isotropy. It is not possible to carry out an analysis in any detail because of the unknown parameters involved, especially in view of the coherent nature of the stripping and compound nucleus processes. It would be interesting to observe the $(d,p\gamma)$ correlation at other deuteron energies and proton angles, in particular at the maximum of the stripping distribution.

V. THE $Be^9(d,\alpha)Li^{7*}(\gamma)Li^7$ REACTION

Following the measurement of the $W(90,\theta_\gamma,180)$ correlation in the $(d,p\gamma)$ process, a measurement of the $(d,\alpha\gamma)$ correlation was obtained with the same experimental arrangement. It was necessary only to change the setting of the gamma channel from 3.4 Mev to 0.5 Mev, and make a small adjustment in the channel of the charged particle detector. With an 0.20-mil Al absorber, (see Fig. 2), the alpha coincidence curve was observed at slightly higher pulse heights than the proton coincidence curve shown in Fig. 2.

TABLE II. Summary of experiments in which no anisotropy has been detected in the emission of the 0.48-Mev gamma ray from Li^{7*}.

Reaction	Correlation	Beam energy, Mev	Ref.
$Li^7(p,p'\gamma)Li^7$	$(p\gamma)$	0.8–1.1	a
$B^{10}(n,\alpha\gamma)Li^7$	$(\alpha\gamma)$	thermal	b
$Li^6(d,p\gamma)Li^7$	$(d,p\gamma)W(0,\theta_\gamma)$	0.7–0.8	c
	$(d,p\gamma)W(90,90,\phi)$	0.63	d
	$(d,p\gamma)W(90,90,\phi)$	0.50	e
	$(d,p\gamma)W(116,\theta_\gamma,180)$	0.41	f
	$(d,p\gamma)W(90,90,\phi)$	0.60	g
	$(d,p\gamma)$	1.00	h
	(d,γ)	0.70	e
$Be^9(d,\alpha\gamma)Li^7$	$(d,\alpha\gamma)W(152,\theta_\gamma,180)$	0.40	i
	$(d,\alpha\gamma)W(90,\theta_\gamma,180)$	0.84	

[a] R. M. Littauer, Proc. Phys. Soc. (London) **A63**, 294 (1950).
[b] B. Rose and A. R. W. Wilson, Phys. Rev. **78**, 68 (1950).
[c] See reference 2.
[d] J. O. Newton, Proc. Phys. Soc. (London) **A64**, 938 (1951).
[e] W. H. Burke and J. R. Risser, Phys. Rev. **87**, 294 (1952).
[f] A. J. Salmon and E. K. Inall, Proc. Phys. Soc. (London) **A66**, 297 (1953).
[g] J. Thirion, Compt. rend. **232**, 2418 (1951).
[h] Phillips, Heydenburg, and Cowie, Phys. Rev. **85**, 742 (1952).
[i] See reference 15.

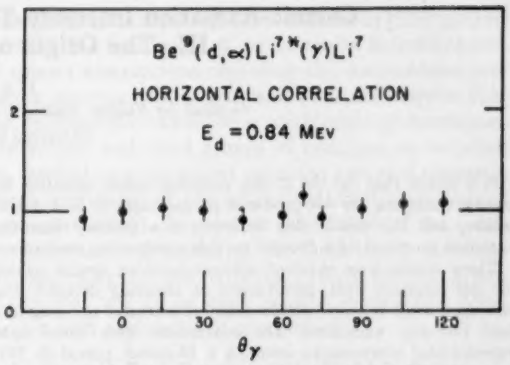

FIG. 8. The $W(90,\theta_\gamma,180)$ correlation for the (d,α,γ) process.

In view of the many experiments, cited in Table II, in which the 0.48-Mev radiation from Li^{7*} is found to be isotropic, it was fully expected to be isotropic in this case also. A measurement of the correlation therefore would provide a good test of the equipment and the method of energy selection. On the other hand, it would give one more "determination" of the spin of Li^{7*}. In this instance a comparison with the markedly anisotropic $(d,p\gamma)$ correlation would be fairly convincing, since both reactions originate in the same manner and involve the same states in the compound nucleus. The $(d,\alpha\gamma)$ correlation has recently been measured by Uebergang and Tanner[15] under different conditions, with an isotropic result.

The results of the present measurement are shown in Fig. 8. It was not thought worthwhile to achieve the same accuracy in this correlation as in the $(d,p\gamma)$ measurements. If the data are folded about $\theta_\gamma=0°$ and 90°, the resulting correlation can be considered isotropic to within about 5 percent. It should be noted that in this measurement the discrimination between the α-γ and p-γ processes is not complete, since the gamma counter, set on 0.5 Mev, will respond to some 3.4-Mev radiation, and the protons and alphas are imperfectly resolved in the particle channel. There is perhaps a trace of the $(d,p\gamma)$ correlation discernible in the $(d,\alpha\gamma)$ measurement in Fig. 8.

A summary of the correlation experiments on the 0.48-Mev gamma ray is given in Table II. There is now abundant evidence for the isotropic distribution of this gamma ray under a variety of conditions in a number of reactions, leading to the assignment of $\frac{1}{2}$ for the spin of the first excited state of Li^7.

We wish to thank Dr. S. M. Shafroth for generous assistance in the early part of this experiment. We have benefited from discussions with Mr. T. S. Schreiber, who is attempting a more quantitative interpretation of the data in this investigation. Mr. Paul Milich aided in the construction and care of the equipment, and Mr. Richard Lee in the operation of the accelerator.

[15] R. G. Uebergang and N. W. Tanner, Australian J. Phys. **6**, 53 (1953).

Cosmic-Radiation Intensity-Time Variations and Their Origin. III. The Origin of 27-Day Variations*

J. A. SIMPSON
Institute for Nuclear Studies, University of Chicago, Chicago, Illinois
(Received October 29, 1953)

It is shown that (a) the 27-day recurring cosmic-radiation intensity variations are not produced by geomagnetic field disturbances, and (b) this 27-day variation is a primary intensity variation produced by a charged particle accelerating mechanism.

These results were obtained without invoking special models for the magnetic field disturbances or requiring detailed time correspondence between cosmic-radiation intensity and magnetic field intensity variations. The conclusions were based upon experimental observations covering a 19-month period in 1951 and 1952. The results are derived from neutron-intensity variations measured as a function of time in aircraft and neutron monitor piles. The data reveal that an indirect association exists between the 27-day intensity variation and geomagnetic disturbances: i.e., geomagnetic disturbances are most likely to occur approximately 2 days after the 27-day maxima of cosmic-ray intensity. It is also shown that intensity changes of ~3–6 percent are sometimes not followed by any geomagnetic disturbance.

An example is given of a nonrecurring sharp intensity decrease of >6 percent, and it is shown that even this event is not produced by geomagnetic field variations.

The results suggest that there is a common mechanism which produces both the accelerating process for cosmic-radiation particles and, indirectly, the geomagnetic disturbances. A search was made for the probable location of such a mechanism. Varying electrical fields of terrestrial origin were considered whereby the incoming primary radiation would undergo an acceleration or deceleration either (1) before entering the geomagnetic field, (2) within the magnetic field region, or (3) after passing through the magnetic field. None of these three possibilities, nor a combined geomagnetic and geoelectric field storm, accounts for all the established experimental facts. In view of these results it is concluded that the accelerating mechanism probably is not of terrestrial origin. The 27-day recurrence corresponds in time to the proper rotation of the solar equatorial latitudes and, since it has been shown that active solar regions at these latitudes are associated with the 27-day cosmic-radiation intensity variations, the required accelerating mechanism is probably controlled by solar processes and may be located near the sun.

From the dependence of the 27-day intensity variation upon particle rigidity, the experimental results show that primary protons undergo the variation, but it is still not proved whether or not particles of $Z \geq 2$ also display this variation.

The experimental data also exclude the production of this variation by the influence of a solar dipole magnetic field.

I. INTRODUCTION

WORLD-WIDE and occasionally large intensity variations of cosmic radiation have been observed over a period of approximately 20 years by using ion chambers and counter telescopes.[1–5] Although these variations usually could be shown to be independent of atmospheric phenomena, it has been widely assumed that these variations were produced directly or indirectly by geomagnetic storms or other geomagnetic field perturbations and, hence, were of terrestrial origin.[2,3,6] In recent years some doubt has been cast upon this assumption since observers have been unable to explain the various aspects of the cosmic-radiation intensity variations on the basis of observed geomagnetic field variations.[6,7–10] It is the purpose of this paper to present experiments which may decide whether or not this assumption is true.

By studying variations of cosmic-ray neutron intensity we have recently shown that it is possible to extend the measurement of intensity-time variations to the low-energy end of the primary-particle spectrum, and in this region we have found intensity variations much larger than heretofore reported for charged particle observations.[11] Since we have shown that meteorological effects do not produce the variations,[12] it is clear either that (a) the prevailing assumption is true and the variations are produced by geomagnetic or geoelectric field perturbations, or that (b) these variations are of primary or extraterrestrial origin, i.e., the intensity variations would be observed even if the earth were removed from the region of the detector, or that (c) some kinds of intensity variations are produced by geomagnetic field variations and other kinds are not.

In earlier preliminary reports we showed that individual examples of intensity variations the order of 5–30 percent in amplitude could not be produced by the geomagnetic field perturbations.[9] In this paper we report the results of a more extensive series of experiments which we believe confirms the earlier results. The study is based upon the largest recurring intensity variations which are to be found throughout the primary-particle momentum spectrum. These are the approximately 27-day recurring variations observed in

* Assisted by the Office of Scientific Research, Air Research and Development Command, U. S. Air Force.

[1] V. F. Hess, Terrestrial Magnetism and Atm. Elec. **41**, 345 (1936).
[2] S. E. Forbush, Phys. Rev. **51**, 1108 (1937); **54**, 975 (1938); S. E. Forbush and I. Lange, Phys. Rev. **76**, 164 (1949).
[3] A. R. Hogg, J. Atm. Terrest. Phys. **1**, 56 (1950).
[4] H. Gheri and R. Steinmaurer, Terrestrial Magnetism and Atm. Elec. **52**, 343 (1947); H. Gheri, Z. Naturforsch. **6a**, 775 (1951).
[5] H. Elliot, *Progress in Cosmic Ray Physics* (Interscience Publishers, New York, 1952), Chapter VIII.
[6] S. Chapman, Nature **140**, 423 (1937).
[7] K. Nagashima, J. Geomag. and Geoelect. **3**, 100 (1951).
[8] D. H. Loughridge and P. F. Gast, Phys. Rev. **57**, 938 (1940).
[9] J. A. Simpson, Phys. Rev. **83**, 1175 (1951).
[10] Neher, Peterson, and Stern, Phys. Rev. **90**, 655 (1953).
[11] Simpson, Fonger, and Wilcox, Phys. Rev. **85**, 366 (1952); Phys. Rev. **87**, 240 (1952).
[12] Simpson, Fonger, and Treiman, Phys. Rev. **90**, 934 (1953); hereafter we refer to this paper as reference I.

the neutron intensity.[11,13] Since it is now established that these variations extend to particle momenta sufficiently high to permit detection with small amplitude by charged particle monitors,[11,13,14] we are confident that the reported ~0.3 percent 27-day variations observed in charged particle detectors from time to time in the past are real and are the same type of recurring variation we have been reporting. The recently found energy dependence of this variation is such that the amplitude of the variation is largest at the low-energy end of the particle differential spectrum.[13] This fact, along with the high counting rates obtained with neutron pile detectors, makes it possible for the first time to study the energy dependence of a single 27-day "cycle" with considerable precision. We find that this kind of variation is composed of moderately well-defined maxima and minima. The neutron observations also show simultaneous series of 27-day recurring variations which appear to be related to active solar regions distributed around the low solar latitudes.[11]

In addition to these recurring intensity variations there are occasional large, sharp decreases of cosmic radiation intensity discovered by Forbush.[2] This is also a world-wide phenomenon. From observations with neutron monitors we find that this type of event usually interferes with the 27-day observations.[4,15] Forbush has associated these occasional large decreases with great geomagnetic storms which do *not* have a 27-day recurrence tendency.

Although no large geoelectric fields have been discovered outside the ionosphere layers, we shall, to consider the origin of these intensity variations more completely, investigate whether or not the 27-day variation could be produced by varying geoelectric fields.

Specifically, we wish to answer the following questions:

(1) Are the cosmic-radiation 27-day recurring intensity variations produced by geomagnetic field perturbations?

(2) Are the intensity variations produced by varying geoelectric fields?

(3) Is the mechanism which produces the 27-day recurrence located at the earth, or is it independent of the earth system but associated with the sun?

II. 27-DAY RECURRING COSMIC-RAY INTENSITY VARIATIONS AND GEOMAGNETIC-FIELD PERTURBATIONS

In order to decide the fundamental question of whether or not geomagnetic-field perturbations produce the 27-day recurring intensity variations, we wish to consider relationships between cosmic-ray and geomagnetic-field variations which are subject to experimental test. These relations may be considered in three general groups. The first group of relationships is independent of special assumptions regarding the mechanisms producing geomagnetic storms and does not require that the geomagnetic storm be a world-wide phenomenon. The second and third groups of relations, as we shall see, depend to some extent upon the assumed magnetic storm mechanism or the assumption that variations of geomagnetic field observed at the surface of the earth also represent variations of the field which interacts with incoming charged particles. Clearly the first group of relations will be of most general interest for comparison with experimental observations.

A. Cosmic-Ray Geomagnetic Field Relations Independent of Special Assumptions

Irrespective of special theories, we know that the geomagnetic field produces the observed "latitude effect" for incoming cosmic radiation. The trajectories of charged particles in a magnetic field are determined uniquely by their directions of arrival and by their ratio, p/Z, of momentum p to charge Z. For a specified rigidity, pc/Ze, certain directions of arrival at the earth may be forbidden. However, at high latitudes we may prescribe, approximately, a unique value for $[p/Z]_\lambda$ at latitude λ above which all rigidities are allowed and below which all rigidities are forbidden. This problem has been discussed in reference I. This is a satisfactory approximation in the latitude range $\lambda > 40°$ for the observations to be reported in this paper.

Since, for continuous observations, we are required to measure the secondary radiations generated by the primaries, we have chosen to observe local neutron production rate as a measure of integrated primary intensity at the top of the atmosphere. In reference I we obtained the relationship between the counting rate R of a neutron detector located at atmospheric depth x, geomagnetic latitude λ, and the vertical, differential primary flux $j_Z(p/Z,t)$ of particles of momentum p and charge Z. We defined the specific yield of neutrons as a function $S_Z(p/Z,x)$ which is experimentally determined from the time-averaged parameters j_z and R to yield the neutron counting rate at depth x arising from a unit flux of vertically incident primary particles of charge Z and rigidity $\propto (p/Z)$. Thus, we found [reference I, Eq. (2)]

$$R_v(\lambda,x,t)=\sum_Z \int_{[p/Z]_\lambda}^{\infty} S_Z\left(\frac{p}{Z},x\right) j_Z\left(\frac{p}{Z},t\right) d\left(\frac{p}{Z}\right), \quad (1)$$

where $[p/Z]_\lambda$ is the cutoff for vertical arrival at λ and $R_v(\lambda,x,t)$ is the counting rate due only to those primaries which arrive from the vertical direction per unit solid angle at time t. R_v and the observed rate R are related in good approximation by the Gross transformation.[16]

[13] W. H. Fonger, Phys. Rev. **91**, 351 (1953); we refer to this paper as reference II.

[14] H. V. Neher and S. E. Forbush, Phys. Rev. **87**, 889 (1952).

[15] A. T. Monk and A. H. Compton, Revs. Modern Phys. **11**, 175 (1939).

[16] S. B. Treiman, thesis, University of Chicago, 1952 (unpublished).

Clearly a variation of R_v may be produced by either a variation of the lower limit of the integral which is determined by parameters of the geomagnetic field, or by a variation of j_s in the integrand, namely, a variation of j_s with time. We shall now consider these two cases in detail.

1. Variations of the Geomagnetic Cutoff

It has been reasonably well established that the primary differential number spectrum for protons at low energies rapidly vanishes near 1 Bev energy (λ approximately 56–58°N).[10,17] For the special case within the atmosphere where the function $S_x(p/Z,x)$ vanishes for finite values of $j(p/Z,t)$ the observed cutoff of the latitude curve is determined by S rather than j and the observed "knee" of the latitude curve will appear at lower latitudes. Sea level ion chambers and counter telescopes are examples of detectors for which $S(p/Z,x)$ vanishes at moderately small values of (p/Z) with the cutoff appearing in the region of $\lambda \approx 45°$.

However, extensive measurements with disintegration product neutron detectors in 1951 and 1952 demonstrate conclusively that $dR/d\lambda \to 0$ near 55° at $x \approx 300$ g-cm^{-2} as we shall later show. Since the cutoff in the primary spectrum is in the range 56°–58°N, we cannot decide at present whether or not $S_x(p/Z,x) \to 0$ for $x > 0$ at the spectrum cutoff. In any case it appears that the cutoff latitudes for the primary spectrum and the observed neutron production near 300 g-cm^{-2}, where the experiments to be reported here were made, are the same within approximately $\pm 2°$ geomagnetic latitude.

The arguments which follow are not dependent upon the origin of these cutoffs.

The integral counting rate R_v is unchanged by variations of $[p/Z]_\lambda$ above the knee of the curve since either $j_s(p/Z,t)$ or $S_s(p/Z,x)$ go to zero above the knee (Changes in the shadow cone above the knee may result from variations of the geomagnetic field; however, this effect is an order of magnitude smaller than the effects reported here.) This effect may be illustrated by using a typical latitude curve and allowing the lower limit of the integral, Eq. (1), to undergo variations $\pm[\delta p/Z]_\lambda$. The typical curve is shown as Fig. 1, curve (*a*). For a variation $+[\delta p/Z]_\lambda$ the integrated intensity will appear approximately as shown in Fig. 1, curve (*b*). For a variation $-[\delta p/Z]_\lambda$ the curve will appear approximately as shown in Fig. 1, curve (*c*).

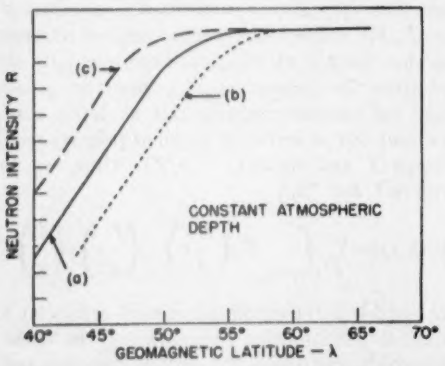

FIG. 1. The predicted behavior of neutron intensity as a function of latitude based upon the assumption that the primary cosmic-radiation intensity variation is produced by a geomagnetic-field variation.

[17] J. Van Allen (private communication).

Thus, perturbations of the geomagnetic field as represented by variations of $[p/Z]_\lambda$ reveal two unique characteristics for the integrated counting rate R_v *vs* λ. *First*, for observations well above the cutoff of the latitude curve, R_v is constant—independent of $[p/Z]_\lambda$ variations. Namely, at $[p/Z]_\lambda$

$$\frac{dR_v}{d\lambda} = -S_s j_s \frac{d[p/Z]_\lambda}{d\lambda} = 0, \text{ where } j_s = 0 \text{ or } S_s = 0. \quad (2)$$

Hence, the fractional change in counting rate is $\delta R_v/R_v = 0$, where $j_s = 0$ or $S_s = 0$. *Second*, the fractional change of counting rate $\delta R_v/R_v$ for $\lambda \gtrsim 35°$ varies as follows:

$$\frac{[\delta R_v/R_v]_{\lambda_2}}{[\delta R_v/R_v]_{\lambda_1}} < 1, \quad \text{for} \quad \lambda_2 > \lambda_1. \quad (3)$$

2. Variations of Primary Cosmic-Radiation Intensity

We now consider the consequences of changes of the differential primary spectrum $j(p/Z,t)$ with time t. From Eq. (1), the maximum value for the integral at time t is obtained by setting the lower limit equal to the minimum particle rigidity $[p/Z]_{\min}$ observed in the primary spectrum. Hence, for all values of $[p/Z]_\lambda < [p/Z]_{\min}$ the integral is a constant. Now, if $j_s(p/Z,t)$ undergoes a variation, the observed counting rate will change for observations at latitudes corresponding to $[p/Z]_\lambda < [p/Z]_{\min}$ as well as for latitudes corresponding to $[p/Z]_\lambda \geq [p/Z]_{\min}$. Thus, an observer within the atmosphere would measure a change in secondary particle intensity above and below the cutoff of the latitude curve. For example, if the fast neutron latitude curve at time t is represented as curve (*a*), Fig. 2, then, if at time t^1 a variation occurs to produce a fractional change of intensity $-\delta R/R$, which variation for simplicity we shall make independent of latitude, then at t^1 the neutron latitude curve will appear as curve (*b*), Fig. 2.

For the special case where the function $S_s(p/Z,x)$ vanishes for finite values of $j(p/Z,t)$ the observed cutoff of the latitude curve will be determined by S rather than j, and again, if j undergoes a variation of intensity with time, the counting rate of the detector will change with time above and below the cutoff determined by S_s.

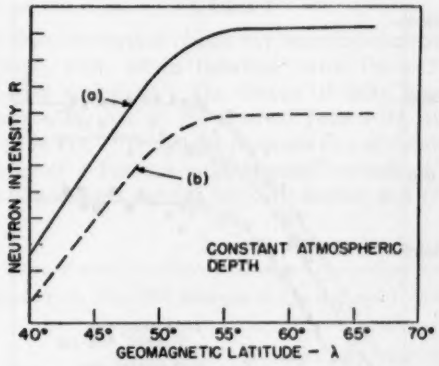

FIG. 2. The predicted behavior of neutron intensity as a function of latitude based upon the assumption that the cosmic-radiation intensity variation is produced by a change in the primary flux with time.

B. Test Experiments

It is clear from the previous discussion that, without invoking special assumptions regarding the origin of geomagnetic field perturbations, we may determine whether intensity variations observed with neutron detectors at high latitudes are produced by geomagnetic field variations or whether the variations are produced by changes in the *primary* intensity. The crucial test consists in determining the behavior of the neutron intensity in the latitude range of $\lambda \sim 40°$ to $\lambda \sim 65°$ during the period of an intensity variation. As noted earlier the largest recurring variations are the approximately 27-day variations in the nucleonic component; hence, we apply the criteria discussed above to the experimental observations of this variation. At constant atmospheric depth a family of latitude curves is to be obtained with each curve representing the different levels of observed cosmic radiation intensity at different times.

We have already reported several preliminary measurements using the detection of fast disintegration product neutrons to measure nucleonic component intensity with B-29 aircraft flying at atmospheric depth $x \approx 300$ g-cm^{-2} in 1948 and 1949. These results demonstrated that observed large variations of cosmic radiation intensity could not be due to geomagnetic field variations.[9] To investigate the behavior of the 27-day recurring variations, a more extensive series of measurements was undertaken in the period 1950–1952 using a type RF-80 jet aircraft. The measurements we report here were obtained in this latter period. The instrumentation was installed in the nose of the aircraft and either atmospheric or local neutron production could be measured with the apparatus. Experimental details regarding the apparatus and flight procedures for the jet aircraft are described elsewhere.[18] No longitude corrections are required for these observations since all data were recorded over a fixed route between $\lambda = 40°$ and 65° N. A single, enriched BF_3 proportional counter surrounded by paraffin with a cadmium shield was used for all the measurements reported in this paper.

The individual curves of neutron intensity *vs* latitude for different times are related by continuous recordings of neutron intensity using pile geometries at selected latitudes. Thus, the observed changes with time of the discrete intensity *vs* latitude curves at high altitudes can be related to long-time continuous intensity variations. Details regarding the continuous neutron monitor stations have been published in reference 1.

Several flights are required to obtain data for a single latitude curve over the latitude range of 40° to 65°N with an aircraft of small range. Because of widely varying weather conditions the sequence of flights may extend over a period the order of 10 to 60 hours. Changes of intensity may occur within this time interval. Consequently, to avoid distortion of a latitude curve constructed from several flights, the data in the latitude interval 40° to 53° (large $dR/d\lambda$) are obtained within a few hours. All additional data required to construct the latitude curve, such as data from flights between $\lambda = 52°$ to 65°, are corrected for the intensity at the time of the lower latitude flights. The individual segments of the latitude curves were fitted by requiring that, for any two completed latitude curves, there must be no discontinuities in the observed change of intensity δR over the entire range of λ. A complete latitude curve constructed from data taken over a period of a few days is referred to a specific day and time, namely, the time for the data obtained in the interval $\lambda \approx 40°$ to 53°. In general, the intensity variations during the collection of data for constructing a single latitude curve were small.

All data reported here are for the atmospheric depth $x = 312$ g cm^{-2}.

1. Case 1

We first consider a case where two latitude curves were obtained at times when *no* magnetic storms were

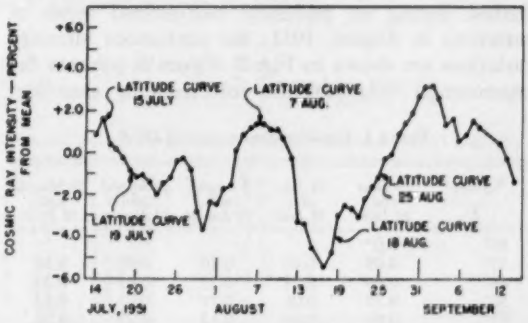

FIG. 3. The percent change of neutron-pile intensity from the mean intensity in pile D-1 using 24-hour average values of intensity.

[18] J. A. Simpson and W. C. Fagot, Phys. Rev. **90**, 1068 (1953).

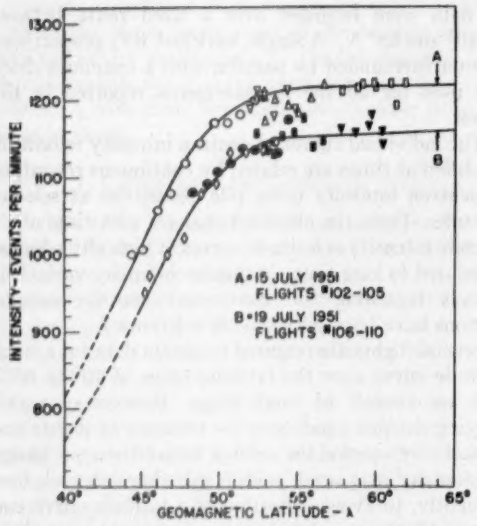

FIG. 4. The neutron-intensity data used to establish the latitude curves for 15 and 19 July, 1951 are shown. The smooth curves (with dashed lines for extrapolations) are used for analysis. The curves are based upon the following aircraft flights (standard deviations are given by the size of the flight identification symbols):

○ 15 July (flight No. 102)
△ 16 July (flight No. 103)
□ 16 July (flight No. 104)
▽ 16 July (flight No. 105)
◇ 19 July (flight No. 106)
● 19 July (flight No. 107)
■ 20 July (flight No. 108)
▼ 20 July (flight No. 109).

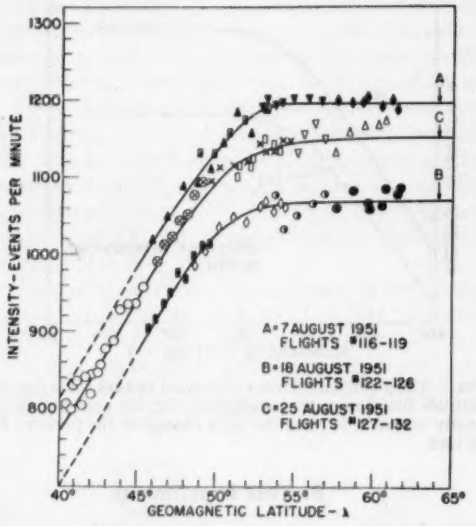

FIG. 5. The neutron-intensity data used to establish the latitude curves for 7, 18, and 25 August, 1951, are shown. The smooth curves (with dashed lines for extrapolations) are used for analysis. The curves are based upon the following aircraft flights (standard deviations are given by the size of the flight identification symbols):

▲ 7 August (flight No. 116)
◪ 7 August (flight No. 117)
▼ 9 August (flight No. 118)
◆ 9 August (flight No. 119)
◑ 17 August (flight No. 122)
● 17 August (flight No. 123)
■ 18 August (flight No. 125)
◇ 18 August (flight No. 126)
▽ 24 August (flight No. 127)
△ 24 August (flight No. 128)
□ 24 August (flight No. 129)
○ 25 August (flight No. 130)
⊗ 25 August (flight No. 131)
✖ 25 August (flight No. 132).

reported during the measurements. The change of daily mean intensity within the interval July 14 through September 9, 1951, is shown in Fig. 3. The first high-altitude measurements were 15–16 July, and we refer these flights to the time of the July 15 flight (see Fig. 4). Similarly, the latitude curve for July 19 and 20 is called the latitude curve for July 19. From these data we have determined the smooth curves which best represent the data for July 15 and July 19. The fractional change of omnidirectional intensity $[\delta R/R]_\lambda$ *versus* λ has been computed in Table I.

2. Case 2

In the following example three latitude curves were studied during an especially well-defined series of variations in August, 1951; the continuous intensity variations are shown in Fig. 3. Figure 5 presents the experimental data and the corresponding smoothed curves. Minor geomagnetic storms were underway during the latter two measuring periods. See Table II, example 8.

Again the evidence supports the conclusion that the observed cosmic-ray intensity variations are due to changes in the primary spectrum. The intensity variations in this period have approximately 27-day recurring variations.[11,13] This is the most complete set of measurements we have been able to obtain within a single 27-day "cycle."

TABLE I. Experimental values of $\delta R/R$.

Date / λ	18 June and 26 June	15 July and 19 July	7 August and 18 August	18 August and 25 August	29 March and 18 June
40°	0.08				
45°	0.08	0.08	0.13	0.08	0.20
48°	0.09	0.07	0.11	0.07	0.22
50°	0.10	0.06	0.11	0.07	0.23
52°	0.09	0.06	0.12	0.07	0.22
56°–65°	0.08	0.06	0.12	0.08	0.22

3. Case 3

In the previous two cases we have considered sequences of measurements within single 27-day intervals. Since the same neutron detector with constant efficiency was used over periods the order of 6–8 months, latitude curves separated by long time intervals may be compared. For example, there is a large difference in observed cosmic-radiation intensity between 29 March, and 18 June, 1951, Fig. 6. The fractional change of intensity $[\delta R/R]_\lambda$ is given in Table I. The measurements for both the March 29 and June 18 curves were obtained during geomagnetic storms.

4. Case 4

A sharp decrease of cosmic-ray intensity was observed in June, 1951, which interfered with the ~27-day recurring variations.[11] The change of daily mean intensity with time in the interval June 4–27, 1951 is shown in Fig. 7. This event possesses two of the characteristics of a Forbush-type decrease,[2] namely, (1) the decrease is rapid and the recovery is slow, and (2) this type of event is infrequent and interferes with the 27-day recurring intensity variations. This decrease is unlike the Forbush-type in that the geomagnetic storm which occurred during the decrease was not unusually large as shown in Table II, examples 6 and 7. Two latitude curves were obtained which are related to this event. The data for June 18 and June 26 latitude curves are displayed in Fig. 8. Clearly, the sharp intensity

TABLE II. Neutron intensity changes and magnetic storms; Climax, $\lambda = 48°$.

Example 1: 21 May 1951 Decrease of *3 percent* starts ~1100 U.T. with no magnetic storm.

								K_w	
20 May	1.8	2.0	1.7	1.6	2.0	1.1	1.7	1.3=13.2	three of the five most quiet days in May
→21	1.0	0.9	0.9	1.1	1.1	1.0	1.4	1.6= 9.0	
22	2.4	1.3	0.6	1.1	2.3	2.3	1.3	1.0=12.3	

Example 2: 23 May 1951 Decrease of *5 percent* starts ~2000 U.T. with no magnetic storm.

								K_w
22 May	2.4	1.3	0.6	1.1	2.3	2.3	1.3	1.0=12.3
→23	2.3	2.1	2.2	3.8	3.6	2.9	5.0	4.1=26.0
24	3.8	2.5	2.0	2.4	2.7	3.4	2.3	2.6=21.7

Example 3: 10 June 1951 Decrease of *4 percent* starts ~2000–2100 U.T. with no magnetic storm.

								K_w	
9 June	3.5	2.6	1.7	2.3	3.1	3.4	0.4	0.8=17.8	
→10	1.0	0.7	0.8	2.4	3.1	2.5	2.1	1.4=14.0	one of five most quiet days, June
11	2.1	2.5	1.7	2.3	2.2	3.1	3.0	2.8=19.7	

Example 4: 29 July 1951 Decrease of *3 percent* starts ~0200 U.T. with no magnetic storm.

								K_w
28 July	3.8	4.4	3.5	4.1	3.2	4.1	4.2	3.1=30.4
→29	2.9	3.4	2.8	3.1	2.7	2.6	2.4	2.2=22.1
30	2.2	2.0	1.8	2.9	2.8	2.7	3.3	3.4=21.1

Example 5: 2 September 1951 Decrease of *3 percent* starts ~1800 U.T. with no magnetic storm.

								K_w	
1 Sept.	2.1	1.6	1.3	1.2	1.9	1.4	0.9	2.6=13.0	four of the five most quiet days in September
→ 2	2.2	1.4	1.1	1.1	1.2	1.4	1.0	1.6=11.0	
3	2.2	1.6	1.9	2.4	2.5	1.2	1.2	2.8=15.8	
4	2.1	2.7	1.4	1.1	2.1	1.9	2.0	2.2=15.5	

Example 6: 14 June 1951 Decrease of *6 percent* starts ~1700 U.T. with magnetic storm s.c. 1750.

								K_w
13 June	2.5	1.6	3.0	2.8	3.4	1.7	1.8	2.2=19.0
→14	0.9	1.1	0.4	1.2	0.9	3.8	4.6	4.6=17.5
15	3.6	2.9	3.3	2.8	3.8	3.4	3.5	3.0=26.3

Example 7: 17 June 1951 Decrease of *3 percent* starts ~1000–1300 U.T. with magnetic storm s.c. 1701.

								K_w	
16 June	3.5	2.3	2.7	2.9	2.7	3.0	2.8	2.5=22.4	
→17	2.3	1.5	1.3	1.0	0.9	4.3	3.6	6.3=21.2	two of five most disturbed days in June
18	5.3	5.2	5.1	3.7	3.3	3.5	2.1	3.2=31.4	

Example 8: 11 August 1951 Decrease of *6 percent* starts ~1700 U.T. and continues for a few days—magnetic storm s.c. 0315 August 11.

									K_w
	10 Aug.	1.6	1.9	2.4	2.3	3.1	1.5	2.0	2.3=17.3[a]
	→11	2.2	2.3	2.5	2.9	3.6	2.9	2.9	3.3=22.6
Note: *4 percent* of total decrease begins ~0200 August 13	12	3.3	2.8	3.1	3.1	3.3	2.8	3.2	3.5=25.1
	→13	2.4	4.6	4.6	3.8	4.4	4.6	4.7	2.2=31.3[b]
	14	1.4	1.2	1.1	2.8	2.4	2.5	2.6	3.1=17.1[a]
	15	3.2	2.2	3.0	1.6	2.5	2.9	4.4	3.3=23.1

[a] Two of the ten most quiet days in August.
[b] One of five most disturbed days in August.

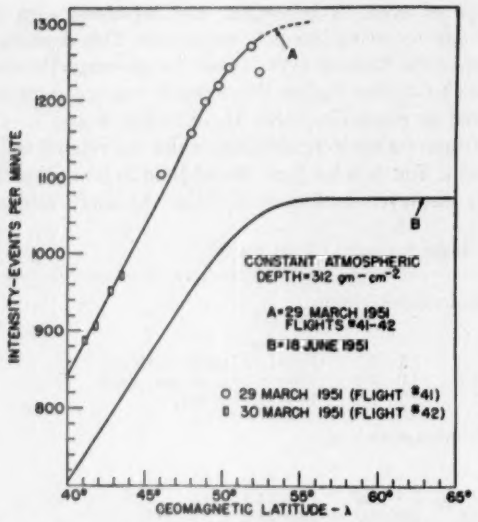

FIG. 6. The neutron-intensity data used to establish the latitude curves for 29 March and 18 June, 1951, are shown. The smooth curves (with dashed lines for extrapolations) are used for analysis. See Fig. 8 for 18 June, 1951, data.

decrease of Fig. 7 was not produced by geomagnetic-field variations.

5. *Further Evidence and Conclusions*

In addition to the above cases selected on the basis of a variety of magnetic storm conditions, we have recorded eight additional latitude curves in 1951–52. To these we add the results of the 1948–49 measurements representing five latitude curves. Without exception, the behavior of these latitude curves is the same as for cases 1–4.

In view of these experimental results we believe *it is now proved that cosmic-radiation 27-day recurring intensity variations are produced by changes of primary-particle flux rather than by variations of the geomagnetic-field intensity.*

C. Search for Association of 27-Day Intensity and Geomagnetic Field Variations

1. *Early Work with Ion Chambers or Counter Telescopes*

In view of the experimental results obtained in the previous section, how may we interpret the careful and extensive work of Forbush, Hess, and others who find evidence indicating much more than an accidental relationship between cosmic-ray intensity variations and changes of geomagnetic-field intensity? We shall now examine this question.

Early studies have often been predicated on the doubtful assumption that a recording magnetometer located at the surface of the earth measures variations of the external horizontal magnetic field component which are proportional to the magnetic field variations in the regions where the field interacts with charged cosmic-ray particles, and that the observed variations of magnetic-field and radiation intensity may be related. If this assumption is valid, there should exist a *detailed time correspondence* between horizontal component field intensity and measured cosmic-ray intensity variations at the surface of the earth. The observations using ion chamber and counter telescope detectors have provided contradictory evidence; however, from these early results there appears to be general experimental agreement that:

(a) large decreases of cosmic-ray intensity may occur with geomagnetic storms;
(b) not all large magnetic storms are accompanied by cosmic-radiation intensity changes;
(c) the magnitudes of the storms do not bear a unique relationship to the magnitudes of the cosmic-ray change;
(d) in some cases there appears to be remarkable time correspondence between the principal variations of the field and cosmic-ray intensity;
(e) those intensity decreases which are associated with geomagnetic storms appear most closely with the main phase of the storm. Subsequently the cosmic-ray intensity slowly returns to "normal" level;
(f) some large cosmic-ray decreases are world-wide phenomena as are the associated geomagnetic storms;[2]
(g) the intensity decreases are observed at high energies (>20 Bev) and down to the cutoff of ion chambers or counter telescopes;
(h) there is a small 27-day recurring cosmic-ray intensity variation. (Independently, it has been established that geomagnetic storms of moderate intensity also possess a 27-day recurrence tendency.)[19]

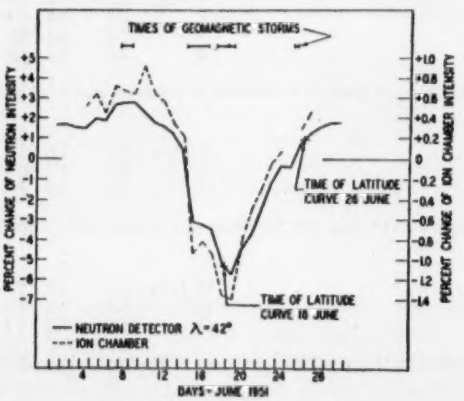

FIG. 7. The percent change of neutron pile intensity from the mean intensity in pile F-1 at $\lambda = 42°$ using 24-hour average values of intensity. The published data from the Sittkus ion chamber at Freiburg, Germany, are shown. The ion chamber percent changes are multiplied by a factor 5 for comparison with neutron intensity changes.

[19] J. Bartels, Terrestrial Magnetism and Atmos. Elec. **40**, 1 (1935). C. Chree, Phil. Mag. Trans. **A213**, 245 (1913).

To account for these observations and the fact that there does not appear to be a unique relationship between cosmic-ray and geomagnetic-field intensity variations, Forbush[3] has invoked a theory of magnetic storms proposed by Chapman.[6] In principle, a variable storm magnetic field produced by an ionic ring current circulating around the earth is added vector-wise to the permanent dipole magnetic field of the earth. The onset of the storm may be sudden (sudden commencement = s.c.) or gradual (no s.c.) followed shortly by the main phase of the storm usually consisting of a decrease of observed horizontal component of the order of $5\times10^2\gamma$ ($\gamma=10^{-5}$ gauss). This model appears to describe satisfactorily many features of the storm.

With a specific model of this type the expected phase relationship and relative magnitude of cosmic-ray intensity and magnetic-field variations have been computed by Forbush,[3] Johnson,[20] Hayakawa,[21] *et al.*, and more recently by Treiman[22] in this laboratory whose results we may briefly summarize as follows:

(1) the *algebraic sign* of the calculated variations $\delta H/H$ is wrong to account for observed variations $\delta R/R$, and
(2) the magnitudes of the cosmic-ray variations are one to two orders of magnitude larger than computed from this model of geomagnetic-field variations.

We conclude from the foregoing evidence that results based on the assumption of detailed time correspondence or on the determination of phase and intensity relationships from a specific model indicate that a relationship between the two phenomena exists but that the measurements are, in some cases, inconclusive or contradictory.

2. *Neutron-Intensity Results*

Owing to these difficulties and to the fact that the 27-day recurring cosmic-ray variation is not produced by geomagnetic-field variations, we approach the problem from a different point of view, using the variations of neutron intensity at high geomagnetic latitudes. This shifts the observations to a lower mean particle momentum range of the primary momentum spectrum. The most readily distinguished feature of the neutron-intensity variations is the time of maximum intensity which has been shown to display the 27-day recurrence phenomenon. We select for study an interval of time containing many of these maxima and determine the most probable time relation of these maxima to well-established features of geomagnetic disturbances, such as periods of greatest field variations, and so forth. Thus, instead of considering the possible

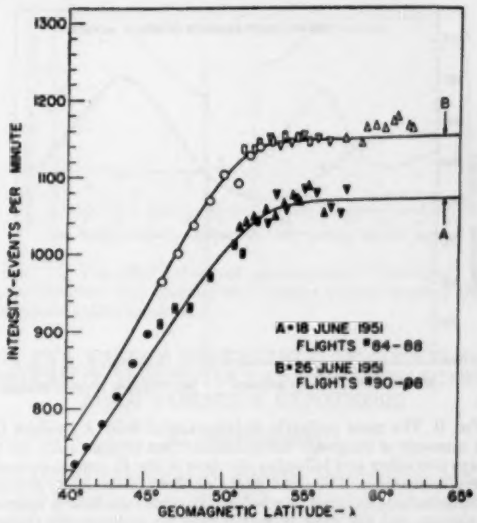

FIG. 8. The neutron-intensity data used to establish the latitude curves for 18 and 26 June, 1951, are shown. The smooth curves (with dashed lines for extrapolations) are used for analysis. The curves are based upon the following aircraft flights (standard deviations are given by the size of the flight identification symbols):

● 18 June (flight No. 84)
■ 18 June (flight No. 85)
▲ 23 June (flight No. 87)
▼ 23 June (flight No. 88)
○ 26 June (flight No. 90)
▽ 30 June (flight No. 96)
△ 30 June (flight No. 97)
□ 30 June (flight No. 98).

geomagnetic-field—cosmic-ray association of individual events, we study the distribution in time of a relatively large number of events to determine the most probable time association between the two parameters being studied.

In a 19-month period of neutron-intensity registration between May 1, 1951, and November 30, 1952, we obtain 41 neutron-intensity maxima that are sufficiently large to be readily identified, without detailed analysis, at two or more widely separated neutron monitoring stations above 40° geomagnetic latitude. Some of the maxima are shown in Fig. 3. In general, the times of these world-wide maxima are defined within approximately ±1 day.

We first investigate how the magnitudes of geomagnetic disturbances are distributed with respect to the times of cosmic-ray intensity maxima. Observers of geomagnetic-field variations have established an arbitrary but reliable measure of the deviations of magnetic-field intensity from average values on undisturbed days. These deviations or disturbances have been reduced to a scale or index of 0 (very "quite") to 9 (extremely disturbed), representing the average deviation of field intensity per 3-hour interval. This scale is called the 3-hour K index. Each magnetic observatory contributes its scaled values to a central group which prepares average K values on a world-wide basis. Thus, the K value for a day consists of an average of eight K values. K_p, the planetary index, is the mean standard-

[20] T. H. Johnson, Terrestrial Magnetism and Atmos. Elec. **43**, 1 (1938).
[21] Hayakawa, Nishimura, Nagata, and Sugiura, J. Sci. Research Inst. (Tokyo) **44**, 121 (1950).
[22] S. B. Treiman, Phys. Rev. **86**, 917 (1952).

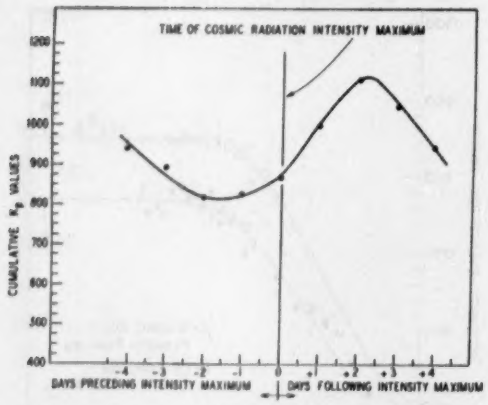
TIME OF COSMIC RADIATION INTENSITY MAXIMUM
CUMULATIVE K_p VALUES
DAYS PRECEDING INTENSITY MAXIMUM
DAYS FOLLOWING INTENSITY MAXIMUM

FIG. 9. The most probable distribution of daily K_p values (K_p is a measure of magnetic disturbances) (see reference 23) for the 4 days preceding and following the days of the 41 cosmic-radiation intensity maxima. The most probable time for smallest geomagnetic disturbances is ~1 day before the cosmic-radiation intensity maximum, and the most probable time for geomagnetic disturbances is ~2 days *after* the intensity maximum.

ized K index from 11 observatories located between geomagnetic latitudes 47 and 63 degrees throughout the world.[23]

Since K_p is a measure of the magnitude of magnetic disturbances, we may determine the most probable distribution of K_p values for the $\pm n$ days with respect to the times of cosmic-ray intensity maxima. Since we do not define the times of maximum closer than one day we shall use the daily sums of K_p indices. After summing for 41 neutron maxima we obtain the results shown in Fig. 9. The largest statistical errors are in the assignment of the times of neutron intensity maximum. However, it is clear that geomagnetic field perturbations increase steadily from $n=-1$ to a maximum at $n=+2$, showing that the times of greatest geomagnetic disturbance will most probably occur 2 days after the appearance of a low-energy cosmic-radiation intensity maximum.

We may check this result by noting that the 10 days each month which have the *smallest* geomagnetic-field disturbances are regularly reported by the same 11 geomagnetic observatories. The distribution of these "quiet" days may be determined with respect to the dates of the 41 neutron-intensity maxima. The results are shown in Fig. 10. A broad but significant maximum appears near $n=-1$. The 5 most disturbed days each month are also reported; an identical analysis shows that a peak lies near $n=+2$ as shown in Fig. 11. The limitation of the number of selected days gives this result a lower statistical weight than the previous results.

[23] For a complete description of K_p see: Geomagnetic Indices *C* and *K*, 1948, Association of Terrestrial Magnetism and Electricity, International Union of Geodesy and Geophysics, Washington, D. C., 1949. For monthly reports of the K_p index see, for example, Series F reports of Central Radio Propagation Laboratory, Natl. Bur. Standards, Washington, D. C.

We can show that the onset of cosmic-ray intensity *decreases*, which follow the intensity maxima described above, is not necessarily accompanied by *any* geomagnetic disturbance. Several cosmic-ray intensity changes of 3 percent or more were selected from a five-month interval of 27-day recurrences which includes the period of the aircraft observations in 1951. The data obtained from the Climax D-1 neutron pile monitor are given in Table II. Values of the K index, K_w, for 36 or more magnetic observatories distributed on a world-wide basis were used to determine the existence of geomagnetic disturbances. From this table we find 5 examples where there is no evidence for magnetic disturbances near the time when the neutron intensity begins to decrease. Several of the magnetic "quiet" days occurred during these periods.

We conclude that large intensity variations need not be accompanied by an observable geomagnetic storm; i.e., $(\delta R/R)/(\delta H/H)\rightarrow\infty$.

A 6 percent decrease is shown in example 6 with close time correspondence between start of the decrease and storm commencement; the K_w values show that only a minor geomagnetic storm occurred.

In view of the already reported relationship[11] between discrete regions of intense solar activity and neutron-intensity maxima these results are of particular interest in interpreting and understanding solar-terrestrial phenomena (for example, with respect to the possible existence of discrete streams of low-velocity ions originating at the sun). However, we shall consider this aspect of the problem in a later paper.

It should be pointed out that, although these relationships may aid in the prediction of geomagnetic storms, more data are required over a greater portion of the 11-year solar half-cycle before these relationships can be considered firmly established.

We now return to the question of the interpretation of ion chamber and counter telescope results. By noting that following a cosmic-radiation intensity maximum an intensity *decrease* certainly occurs, we

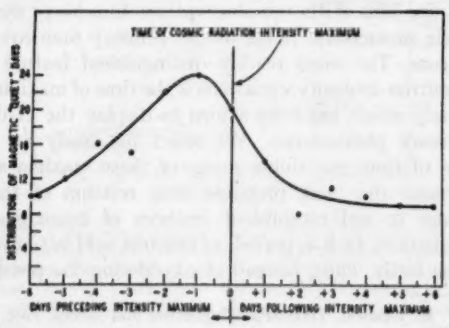
TIME OF COSMIC RADIATION INTENSITY MAXIMUM
DISTRIBUTION OF MAGNETIC "QUIET" DAYS
DAYS PRECEDING INTENSITY MAXIMUM
DAYS FOLLOWING INTENSITY MAXIMUM

FIG. 10. The distribution of geomagnetic "quiet" days (10 are reported each month) with respect to the times of the 41 neutron-intensity maxima.

may restate our neutron intensity observations as follows:

(1) the most probable time for the commencement of a geomagnetic storm is near the beginning of a cosmic-radiation intensity *decrease*, and
(2) the most probable time for geomagnetic disturbances is during a decrease of cosmic-ray intensity.

Thus, the properties outlined in 1(a–h) of this section for ion chambers may be interpreted from the point of view of the neutron intensity observations. Whether or not the large and rare great magnetic storms and cosmic-ray intensity decreases—the Forbush-type decreases (see Sec. B, case 4)—also possess these properties is not known.

D. On The Location of the Phenomenon Associating Cosmic-Ray and Magnetic Field Variations

The nature of the indirect link or association between cosmic-ray variations and magnetic-field disturbances is obscure; however, there are two limiting possibilities:

(a) the link occurs in the region of the earth and is of terrestrial origin;
(b) the link is independent of the earth system and is of solar origin.

To distinguish between these alternatives we noted that the variation of counting rate R in Eq. (1) must arise from changes in $j_s(p/Z,t)$. Since this differential number spectrum is a function of p/Z, a change δj may occur either through a change in the number of particles in the primary spectrum, *or* through an acceleration (or deceleration) of particles which reach the earth. Consequently, the problem of identifying the location of the cosmic-ray—magnetic-field link reduces to the following alternatives; either

(1) the cosmic-radiation intensity variations are produced by *geoelectric* fields and are, therefore, of terrestrial origin, or
(2) the intensity variations are of extraterrestrial origin and are produced by injection, acceleration, or perturbation processes (see Fig. 1 of reference I).

If (1) is correct, the geoelectric fields are probably related to the geomagnetic field disturbances; if (2) is correct, the cosmic-radiation intensity variations are of primary origin and are indirectly associated with terrestrial magnetic field perturbations by means of a solar-controlled mechanism outside the earth system. Although at present geoelectric fields are not known to exist outside the ionosphere, we shall consider this question in the next section, using the results from measurements of nucleonic component intensity-time variations.

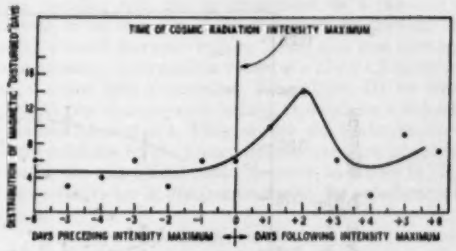
TIME OF COSMIC RADIATION INTENSITY MAXIMUM
DISTRIBUTION OF MAGNETIC "DISTURBED" DAYS
DAYS PRECEDING INTENSITY MAXIMUM
DAYS FOLLOWING INTENSITY MAXIMUM

FIG. 11. The distribution of geomagnetic "disturbed" days (5 are reported each month) with respect to the times of the 41 neutron-intensity maxima.

III. THE ENERGY DEPENDENCE OF COSMIC-RAY INTENSITY VARIATIONS AND THE GEOELECTRIC FIELD VARIATION HYPOTHESIS

A. Introduction

In this section we shall consider observations related to the question: are the 27-day recurring cosmic-ray intensity variations produced by varying geoelectric fields? This is a difficult question to evaluate, particularly since there is no experimental evidence to show that geoelectric fields exist outside the quasi-equipotential "surface" of the terrestrial ionosphere. The question of the existence of distant varying electric fields is deferred for later discussion.

Without requiring special models for the production of varying electric fields in the region of the earth we may note that cosmic-ray intensity variations arise from three general factors whenever such an accelerating-field variation is operative:

(1) there is a contribution to the average intensity since the number spectrum is a function of energy; i.e., $dj(E,t)/dE \neq 0$;
(2) there is a change in the number of particles found in an energy interval dE which arises from the application of Liouville's theorem. This was first pointed out by Nagashima.[7]
(3) If the particle acceleration occurs within the geomagnetic field, a contribution to the change of intensity is produced by changes of the observed cutoff rigidity at the earth.

Thus, whenever a geoelectric field undergoes a variation, the accelerated charged particles will form an energy spectrum which differs from the mean differential energy spectrum, and it is this fact which offers the possibility of comparing the effects of an assumed varying electric field with experimental observations.

Three properties of cosmic-ray intensity variations are accessible for experimental observations if varying geoelectric fields produce the intensity variations. They are:

(1) The magnitudes of the intensity variations are dependent on particle energy. The detailed form of the energy dependence is determined by the

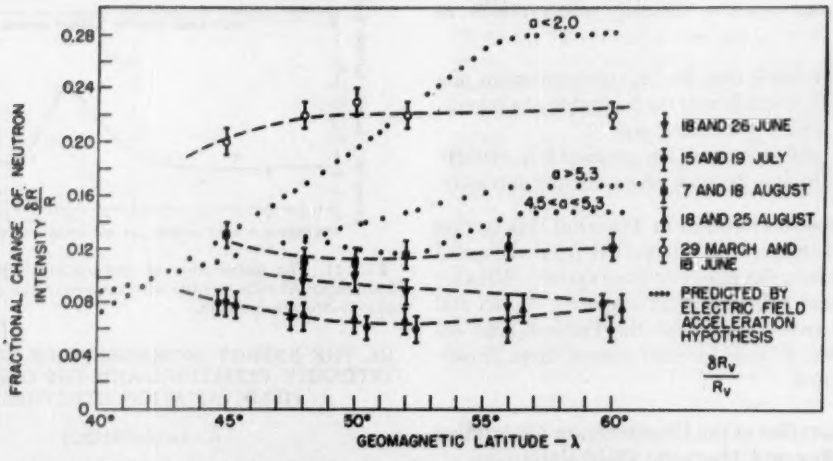

Fig. 12. The experimentally determined ratio $\delta R/R$ is shown with experimental errors for the latitude range 45°–60°. The calculated ratio $\delta R_v/R_v$ is shown for three cases where a geoelectric field is assumed to produce the required particle acceleration. Preacceleration corresponds to $a>5.3$. Mixed field acceleration corresponds to $4.5<a<5.3$. Postacceleration corresponds to $a<2.0$. These calculated curves are normalized to 0.10 at $\lambda=42°$ for comparison with the experimental results.

assumed properties of the geoelectric field model as we shall indicate later.

(2) From Eq. (1), Sec. II it is clear that intensity variations are predicted at geomagnetic latitudes, not only where $dj/dE\neq 0$, but also where $dj/dE=0$ (above the knee of the integral geomagnetic latitude curve). This is contrary to the results from perturbations due to geomagnetic field cutoff.

(3) All charged particles, $Z=1,2,\cdots$, experience the common accelerating mechanism.

Whereas (3) requires detailed measurements at the top of the atmosphere, (1) and (2) may be studied within the atmosphere. We demonstrated in reference I that the neutron intensity variation method is most suitable for the study of energy dependence of intensity variations Hence, the aforementioned properties (1) and (2) are measured using neutron detectors within the atmosphere at intermediate and high geomagnetic latitudes.

B. Experimental Observations

The neutron intensity measurements obtained with aircraft and neutron piles at fixed locations which were described in Sec. II will also be used here. It is most convenient to compare experiment with theory by comparing the observed fractional changes of counting rate $\delta R/R$ (as defined in Sec. II for the omnidirectional detector) with the ratio $\delta R/R$ calculated from specific models for geoelectric field variations. Since the calculations are carried out for vertically incident primary particles of charge Ze at the top of the atmosphere, it is the fractional change arising from these vertically incident particles which is calculated, namely $\delta R_v/R_v$. But since the absorption mean free path for the neutron-producing radiation L is a function of λ at small atmospheric depths, we have for the region below the air transition-curve maximum the relation (reference I):

$$\frac{\delta R_v}{R_v}=\frac{\delta R}{R}-\frac{\delta L}{L}\left[\frac{x/L}{1+x/L}\right].$$

For our present aircraft measurements of $\delta R/R$ at $x=312$ g cm^{-2}, we must determine the importance of the second term in δL in the latitude range 40°–65° before the experimental results can be compared with computed values of $\delta R_v/R_v$. As we shall observe subsequently from the hypothesis of varying electric fields, we only require the sign and order of magnitude of this term. Wherever $j_v(p/z,t)=0$ or $S_v(p/z,x)=0$, we know that $L(\lambda)=$ constant and

$$\delta R_v/R_v=\delta R/R.$$

Elsewhere, in the region $56°>\lambda>40°$ an increase $+\delta R$ produces a change in mean free path $-\delta L$. Hence, for an increase (or decrease) of intensity in general, we have

$$\delta R_v/R_v\gtrless\delta R/R.$$

For the measurements reported here, the range of L values is 155 to 180 g cm^{-2} and the contribution of the term δL is small.

The experimental values of $\delta R/R$ are given in Table I and are plotted as a function of latitude λ in Fig. 12. As a consequence of the above arguments on the effect of δL *vs* λ, these experimental curves represent the *lower limits* for the fractional changes of counting rate for $\lambda<56°$.

C. Assumed Geoelectric Fields

Without postulating at this time the mechanism whereby a geoelectric field is produced, we may determine its effect upon

cosmic-ray intensity if the radial extension or spatial limits of the field are specified. There are three cases we wish to consider: (1) The electric field accelerates (or decelerates) the primary charged cosmic radiation prior to its being deflected in the geomagnetic field—we call this case *preacceleration*; (2) the cosmic rays are accelerated by an electric field within the geomagnetic field—called *mixed field acceleration*; (3) the cosmic rays are accelerated by electric fields after passing through the geomagnetic field—called *postacceleration*.

1. Preacceleration

The expected intensity variations for charged particles accelerated by a varying electric field prior to their deflection in the geomagnetic field may be calculated with the following assumptions:

(a) The particles arrive vertically incident at the atmosphere.
(b) Both protons and heavy nuclei are accelerated and the yield of disintegration product neutrons per primary particle nucleon is independent of the atomic weight of the primary particle (see reference I).
(c) The primary spectrum at "infinity" is known and is given by the time averaged spectrum at the top of the atmosphere.

With these assumptions the expected dependence of the variation $\delta j(p/z,t)$ upon magnetic rigidity may be determined from reference II. Using the neutron specific yield as a function of λ, the dependence of $\delta R_s/R_s$ upon geomagnetic latitude has been calculated for $x=312$ g cm^{-2} for the latitude range $\lambda>40°$ (see Table III and Fig. 12). It is clear that the calculated latitude

TABLE III. Values of $\delta R_s/R_s$. $x=312$ g cm^{-2}.

λ	$a<2.0$	$3.6<a<4.5$	$4.5<a<5.3$	$5.3<a$
42°	1.00	1.00	1.00	1.00
48°	1.74	0.91	1.29	1.29
52°	2.24	1.18	1.18	1.49
56°	2.82	1.48	1.48	1.67

effect for $\delta R_s/R_s$ between $\lambda=40°$ and $\lambda<56°$ is

$$\frac{\delta R_s}{R_s}(56°)\Big/\frac{\delta R_s}{R_s}(40°)=1.7,$$

whereas from the experimental points in Fig. 12 we note that this ratio must be <1, since, for $\lambda<56°$, $\delta R_s/R_s>\delta R/R$ as shown earlier.

2. Mixed Field Acceleration

For preacceleration the geomagnetic cutoff rigidity of charged particles is unchanged, i.e., the changes in observed intensity are due only to $\delta j_s(p/z,t)$. However, for cases of particle acceleration within the geomagnetic field there is an additional contribution, namely, the lower limit of Eq. (1), undergoes a change $[\delta p/Z]_\lambda$ at the earth. For the cases of mixed field acceleration and postacceleration this latter contribution has been computed by Treiman and Jory[24] in this laboratory. Without requiring a detailed model for the production of the electric field, their calculations are based upon the following additional assumptions:

(a) The geoelectric field is assumed to have axial symmetry about the magnetic axis of the earth.
(b) In the equatorial plane the radial extent of the electric field, Δr, is small compared with the distance of the field from the axis of the dipole geomagnetic field, i.e., $\Delta r \ll ar_s$, where r_s is the radius of the earth.

If ar_s is the distance between the geomagnetic dipole and the region Δr of strong electric field, then the fractional change in neutron counting rate can be determined as a function of λ for various values of a. For $a>5.3$ the cosmic-ray particles are in the outer allowed Stoermer region; hence, this case corresponds to preacceleration. Intermediate values of a ($2<a<5.3$) are in the region of mixed field acceleration. From Table III we note for $3.6<a<5.3$ that there are some latitude intervals for which $\delta R_s/R_s$ is an inverse function of λ. These are the short latitude intervals where the solutions for the preacceleration cases are joined to the solution for the mixed field cases. However, as shown in Fig. 12, even these results are in disagreement with the experimental observations.

3. Postacceleration

For $a\sim1$ we obtain the limiting case where the charged particles are accelerated in a varying electric field after their deflection in the geomagnetic field. These conditions can only be approximately satisfied near the surface of the earth.[25] For this case the calculated dependence of $\delta R_s/R_s$ on λ is also large and in disagreement with observations.

D. Other Evidence

In the preceding discussion we have shown the substantial disagreement at the low-rigidity end of the primary spectrum between experimental observations and the results predicted by assuming the existence of varying geoelectric fields. On the other hand, by using the experimental data at mountain and aircraft altitudes, given in Sec. II, we can show that the fractional variation of total intensity, $\delta R/R$, is a function of atmospheric depth x. Therefore, it cannot be independent of particle energy for all particle energies. For example, if we consider the dependence of $\delta R/R$ on x at $x=48°$, we find that the ratio of $\delta R/R$ ($x=312$ g cm^{-2}) to $\delta R/R$ ($x=680$ g cm^{-2}) is 1.7 ± 0.2 using the data from the D-1 monitor. In reference II it was shown that the 27-day variation is energy-dependent in the mean energy ranges of $\sim$7–40 Bev for protons and that this is consistent with the energy dependence required by a preacceleration-type electric field. We encounter the difficulty that none of the forms of energy dependence postulated thus far explain all the experimental data, and we cannot identify the location of the accelerating mechanism on the basis of energy dependence alone. Hence, we shall investigate the extent to which the other known properties of this variation may assist in resolving this problem.

There are several well-established properties of the 27-day intensity variation. They are summarized as follows:

(a) The variation is not produced by meteorological effects.
(b) The variations are recurring and not periodic. The variation may extend to as many as 9–11 cycles or to as few as 2 or 3 cycles.
(c) During the progress of a 27-day sequence of variations, a second sequence is sometimes present. Hence, at least two series of recurrences may occur simultaneously.

[24] S. B. Treiman and F. Jory (unpublished).

[25] The potential between the ionosphere and the surface of the earth is the order of 4×10^5 volts; hence, even for the primary particles of lowest observed rigidity an acceleration by this field is only $a\sim0.01$ percent amplitude effect and, therefore, negligible.

(d) The variation is a world-wide phenomenon.
(e) The variation is not produced by geomagnetic field variations.
(f) In a single 27-day "cycle" the maximum to minimum intensity variation can be as large as ~20 percent–30 percent by including the low-energy portion of the particle spectrum.
(g) The *average* maximum to minimum intensity variation for a single cycle has been decreasing in successive years from 1951 through 1952 and into 1953.
(h) Especially at low primary-particle energy, the variation is composed of a series of irregular maxima and minima. There is no evidence for a constant level of intensity interrupted by intensity decreases at 27-day intervals.
(i) The ratio of the intensity variation between a particle detector observing a proton spectrum of mean energy ~7 Bev and a detector observing a proton spectrum of mean energy ~40 Bev is 5:1. Thus, the 27-day variation displays an energy dependence, and particles of the order of 40-Bev energy are observed to undergo this variation.
(j) At very low primary-particle energies the variation has a negligible energy dependence.

To proceed with the discussion, taking into account the foregoing facts, we introduce the reasonable assumption that world-wide magnetic and electric field variations and storms are manifestations of a common electromagnetic system; i.e., geoelectric field storms coexist with geomagnetic field storms. We do not specify the nature of the moving charges or their distribution except to note that the current flow must be in a westward direction to account for the phase of well-established features of geomagnetic storms.

We now examine the foregoing (a) to (j) properties insofar as they relate to the identification of the accelerating mechanism which produces the 27-day variation. Property (a) was discussed extensively in reference I and will not be considered further here. Properties (b) and (c) are common to either a solar or a terrestrial phenomenon induced by solar processes; hence, they provide no unique identification of the origin of the variation. Properties (d) and (e) likewise are indecisive since, by invoking suitable assumptions, either alternative could be supported.

Since the intensity variation is world-wide, it is clear that property (f) implies that a large fraction of the total *number* of particles in the cosmic radiation incident on the earth undergoes this variation. Thus, from the point of view of the solid angle involved it would be most desirable to locate the mechanism close to and surrounding the earth. To fulfill this requirement the magnitude of the electric field variations would be $\sim 4 \times 10^8$ volts or more to account for a 20 percent variation. (Variations of at least 30 percent were observed, but, although they are produced by an accelerating mechanism, we have not proved that these variations are part of a 27-day sequence.[9]) The geoelectric field hypothesis in its present form fails to satisfy this condition since the magnitude of electric fields generated by a wide range of storm current systems is at least an order of magnitude less than this value.[7]

A further difficulty with the assumption of large, varying geoelectric fields during geomagnetic storms arises from the evidence given in Table II wherein 3–6 percent changes of cosmic-radiation intensity are unaccompanied by any observable geomagnetic storm and, hence, by any large geoelectric field variation. On the other hand, it is known that large magnetic storms may produce no change of cosmic-radiation intensity. Thus, it is difficult to support any theory which involves electromagnetic field variations derived from circulation currents in the region of the earth.

Continuous recording of neutron intensity variations with 3 or more widely separated detectors began in 1951 and the recordings clearly show that in the subsequent period of time the average amplitude of the 27-day intensity variation has been decreasing (property (g)). However, we have no evidence to show that the magnitudes of the geomagnetic storms in the same period have decreased.

In an attempt to explain the large Forbush-type sharp decreases Nagashima has invoked the geoelectric field variations which accompany geomagnetic storms. The required westward current flow which occurs during magnetic storms restricts the production of geoelectric storm fields to decelerating fields. Hence, only intensity decreases are expected during magnetic storms. In view of property (h), however, it is clear that these arguments cannot be introduced at present as evidence for or against the production of the 27-day variation by geoelectric field storms.

We have already presented considerable evidence to show that properties (i) and (j) require a varying acceleration mechanism to explain the observed 27-day recurring variations. In view of the high-energy particles which undergo this variation an accelerating mechanism which could modulate an already existent primary particle distribution would be the simplest assumption to make tentatively at the present time.

E. Charged Earth Hypothesis

The cosmic radiations bring to the earth sufficient positive charge to produce 10^9 volts potential with respect to infinity in the order of 30 days if the earth is assumed to be a spherical, equipotential surface that can retain the accumulated charge. This time for charge buildup would be even less if charged particles in the auroral regions are included. However, Ferraro[26] has considered this problem. He finds that to maintain an ionosphere the upper limit for the accumulated charge would provide a decelerating potential of less than 10^8

[26] V. C. A. Ferraro (private communication).

volts for incoming charged particles. Thus, no Coulomb-type field accounts for the acceleration mechanism.

F. Conclusions Regarding a Geoelectric Field Hypothesis

There are serious difficulties in accounting for the 27-day variation by a geoelectric field. No proof has been given that such a field does not exist. Rather, the foregoing evidence indicates that even if a nearby electric field exists it does not by itself or in combination with a geomagnetic field satisfactorily account for the observed properties of the 27-day cosmic-radiation variations.

In view of these conclusions we are led to the assumptions that the cosmic-ray accelerating mechanism is of extraterrestrial origin, and that the association or link between processes on the sun and this accelerating mechanism is independent of the earth system.

IV. THE 27-DAY VARIATION AND THE GENERAL SOLAR MAGNETIC FIELD

Many investigators have searched for a periodic 27-day intensity variation which could be explained by a solar eccentric dipole field. Vallarta and Godart[27] have computed the effect of the solar plus Stoermer cones on the cosmic-radiation intensity as a function of time and predicted a small 27-day periodicity. It might be argued that the recurring variation discussed in this paper could, indeed, be periodic with occasional periods concealed by an unknown interfering intensity variation.

We can show that this argument cannot be accepted for the following reasons:

(1) The solar-terrestrial cone does not influence particles at the high rigidities which are observed to undergo the 27-day variation.
(2) At low primary-particle rigidity the magnitude of the observed variation is an order of magnitude greater than for a solar-cone induced variation.
(3) A single 27-day period is predicted, whereas simultaneously at least two 27-day sequences are sometimes observed.
(4) Since the observed 27-day period is the *synodic* rotation period, we may compute the proper solar rotation from this observation. Let us assume that the observations permitted a synodic period as long as 28 days; then the observed proper solar rotation becomes 26 days. Since the proper rotation of the sun changes from $\sim$25.2 days near the equator to over 30 days near the poles, it is obvious that the region of the sun related to the cosmic-ray variations, which we have reported in this paper, lies in the equatorial zone and not in the polar regions. Hence, the period predicted by a solar cone is in disagreement with observations unless the solar magnetic poles lie near the equator.

[27] M. S. Vallarta and O. Godart, Revs. Modern Phys. **11**, 180 (1939).

It is clear from these arguments that the observed 27-day recurrence intensity variation described by neutron and ion chamber intensity variations is not the phenomenon predicted by Vallarta.

If a periodic variation exists, it must be small and appear at low primary-particle energies and would probably be covered up by the much larger 27-day recurrence variation. However, in view of the increasing evidence that (a) the general magnetic field of the sun is small and (b) that whatever field does exist may be of a higher order than a dipole, the existence of a measurable, true "27"-day periodicity is unlikely.

V. KINDS OF PARTICLES PRODUCING THE 27-DAY VARIATIONS

It was shown in reference I that the specific yield of neutrons for primary particles of $Z \geqslant 2$ decreases to zero for $\lambda \gtrsim 45°$. Since the 27-day intensity variation is observed at latitudes up to the cutoff near $\lambda = 55°$, it is clear that primary particles of charge $Z = 1$, i.e., protons, are undergoing the variation. Whether or not particles of $Z \geqslant 2$ also contribute is an open question which would provide a critical test for the identification of the accelerating mechanism.

CONCLUSIONS

We have shown that cosmic-radiation intensity variations possessing a 27-day recurrence are not produced by geomagnetic field variations. On the other hand, since we find an indirect association between the intensity maxima of these 27-day variations and geomagnetic field disturbances, we have investigated whether there is a mechanism common to both processes, and whether the association is of terrestrial origin. The experimental observations show that the variations are due to changes in the primary particle intensity and, therefore, must be produced by a charged particle accelerating mechanism. We have investigated the possibility of a nearby accelerating-decelerating electric field having a 27-day recurrence. Neither a geoelectric field alone nor combined with geomagnetic field variations fit the experimental facts, and we conclude, on the basis of our present experimental knowledge, that the acceleration mechanism may be of extraterrestrial origin.

Since we find that a 27-day recurrence observed in the earth coordinates is consistent with the proper motion of the solar equatorial band of latitudes, and since it is known that special solar regions located in this solar latitude range are associated with both geomagnetic storms and the cosmic-ray 27-day variation, we conclude that the mechanism which changes the energy distribution of the primary-particle spectrum is controlled by processes on the sun.

We have also shown that primary protons undergo the 27-day intensity variations, but the extent to which particles of $Z \geqslant 2$ participate remains unknown.

An example of a nonrecurring, sharp intensity decrease of >6 percent was found which may be the type

discovered by Forbush. It is clear from the observations that this large intensity decrease was not produced by a geomagnetic storm.

It may develop that the solar-terrestrial associations indicated by these neutron intensity measurements will find application in interpreting and predicting the occurrence of solar related phenomena.

The author is particularly indebted to the pilots and ground crews of the RF80 No. 8430 and the Flight Test Division of the Wright Air Development Center for making it possible to accomplish the many difficult flights required for these measurements. The author also appreciates the aid of L. Wilcox, A. Vernon, K. Benford, and R. Baron in preparing the flights and data; and of P. Shevick and N. Wood in contributing to the development of the airborne circuits and counters, respectively. The helpful criticisms of Dr. J. W. Firor along with the calculations by Dr. S. B. Treiman and Mr. F. Jory on the effects due to mixed electric and magnetic field variations were deeply appreciated.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

The Origin of Cosmic Rays*

PHILIP MORRISON, *Cornell University, Ithaca, New York*

AND

STANISLAW OLBERT AND BRUNO ROSSI, *Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received December 23, 1953)

The isotropy and composition of the primary cosmic radiation suggest that cosmic rays are trapped within the galaxy for an average time of the order of 10^6 years,—a long time compared with the time of escape along straight-line paths, but short compared with the mean life against nuclear collisions with interstellar matter. If one accepts this conclusion, it appears possible to account for the observed properties of cosmic rays under the assumption that cosmic rays acquire their large energies through a gradual acceleration in space, such as suggested by Fermi. In contrast to the original Fermi theory (which denied any possibility of escape from the galaxy), we now find that the energy spectra of protons and heavier nuclei are approximately the same, and that the required injection energies are very modest for all components. We are obliged, however, to assume a much faster rate of acceleration than the original theory required.

In this paper we develop in some detail the consequences of the above assumptions on the basis of a specific model, describing the motion of cosmic rays through the galaxy as a random motion between scattering centers represented by moving magnetized clouds. We briefly discuss the astrophysical implications of our assumptions and the plausibility of the model.

I. GENERAL CONSIDERATIONS

A. Introduction

FERMI[1] has proposed a theory of the origin of cosmic rays according to which the cosmic-ray particles diffuse randomly in interstellar space, gaining energy by collisions against moving magnetic fields, until eventually they loose their accumulated energy catastrophically, by collisions with hydrogen nuclei. This theory explains in a natural way the general isotropy and the observed energy spectrum of the cosmic-ray protons. It fails to account satisfactorily for the considerable flux of alpha particles, and for the heavier nuclei in the primary cosmic radiation. The difficulty is twofold. In the first place, according to Fermi's theory, the injection energy, i.e., the energy required for initiating the acceleration process (an energy at which the rate of energy gain overtakes the rate of loss of energy by ionization), is extremely high for the heavier components. In the second place, the energy spectrum computed for the heavier particles falls off much more steeply at high energy than that of the protons, because their mean free path against collisions with hydrogen nuclei is much shorter. Experimentally, however, the energy spectrum of the various components seem quite similar up to some 10^{13} ev per nucleon at least.[2]

Both difficulties can be overcome if one assumes that cosmic-ray particles diffuse around the galaxy for a time long compared with the average time for escape from the galaxy along straight-line paths, yet short compared with the mean life before collisions with interstellar hydrogen. Under this assumption (which, as originally pointed out by Bradt and Peters, has strong experimental support[3]) the mean life of a cosmic-ray particle in the galaxy is determined mainly by the escape probability, and is thus roughly independent of its mean free path for nuclear collisions. Then the

* Supported in part by the joint program of the U. S. Office of Naval Research and the U. S. Atomic Energy Commission.

[1] E. Fermi, Phys. Rev. 75, 1169 (1949).

[2] Kaplon, Peters, Reynolds, and Ritson, Phys. Rev. 85, 295 (1952).

[3] H. Bradt and B. Peters, Phys. Rev. 80, 993 (1950); see also B. Peters in *Progress of Cosmic Ray Physics*, edited by J. G. Wilson (North-Holland Publishing Company, Amsterdam, 1952).

energy spectra of the various components are approximately the same. The mean life of a cosmic-ray proton in the galaxy is much shorter, according to the present suggestion, than on the original Fermi theory. To account for the observed spectrum, it therefore becomes necessary to assume a rate of energy gain by collisions with moving magnetic fields considerably greater than that assumed by Fermi. The high rate of energy gain decreases the required injection energy to very plausible values, not only for protons, but even more strikingly for the heavy components.

However, it is doubtful whether or not the high rate of energy gain in compatible with astrophysical evidence. We do not intend here to minimize this difficulty. Indeed, we wish it to be clearly understood that the main purpose of this paper is not to uphold a given theory of the origin of cosmic rays, but rather to specify the conditions that must be met by any theory that explains the high energies of cosmic-ray particles by a mechanism of gradual acceleration in the motion through space.

B. Mean Life of Cosmic Rays in the Galaxy

We consider the relevant part of the galaxy as a thin, circular, disk-shaped volume, with negligible azimuthal variations. This shape is the shape observed for the gas clouds, the dust, and the hot stars of the galactic Population I. The radius of the disk is R_0, and its thickness $2h$.

From astronomical data we take the following approximate dimensions for this galactic disk:

$$h=10^3 \text{ light years}, \quad R_0=5\times10^4 \text{ light years},$$

and for the position of the sun:

$$z_s=30 \text{ light years}, \quad R_s=3\times10^4 \text{ light years}.$$

These are generally accepted figures, but no great accuracy can be assigned to them. The most important figure, that for h, is not very well defined.

We assume that interstellar matter consists entirely of hydrogen, spread through space with a mean density of $\frac{1}{2}$ proton per cm^3. We can estimate the mean free paths, l, for nuclear collisions in this interstellar space for the various components of the cosmic-ray beam, using a value of 4×10^{-26} cm^2 for the nucleon-nucleon cross section at cosmic-ray energies and computing the cross sections of the heavier components on the basis of the semitransparent model of nuclei. The mean free path in nuclear matter is taken as 2.7×10^{-13} cm, and the nuclear radius as $1.4\times10^{-13}A^{\frac{1}{3}}$ cm. Table I gives the results.[4]

[4] Of course some of the collisions will result only in a small change of the atomic number in the case of complex nuclei, or in a small change of the kinetic energy in the case of protons. Therefore the *effective* mean free paths may be appreciably larger than the actual collision mean free paths. However, our rudimentary knowledge concerning the density of interstellar matter, the cross sections for high-energy collisions and the events following such collisions, does not justify a more detailed evaluation of l.

TABLE I. Mean free paths, l, for nuclear collisions (density of interstellar hydrogen: $\frac{1}{2}$ proton per cm^3).

Cosmic-ray particle:	Proton	α particle	CNO group	Fe group
Average atomic weight	1	4	14	30
l (measured in 10^3 l.y.)*	50	20	7	4

* "l.y." = light years.

It is important to remark that, irrespective of the acceleration mechanism, any galactic theory of the origin must ascribe to cosmic-ray particles a mean free path before escape from the galaxy, L, which is long compared with the thickness of the galactic disk, but short compared with the collision mean free path. The upper boundary to L comes from the mass spectrum of the cosmic-ray beam. Considerations based on the relative abundance of protons and heavier nuclei in this beam, and on the numbers of secondary protons which a pure heavy primary beam would make by nuclear collisions with interstellar matter, show that the mean path L of cosmic-ray particles (at least up to 10^{13} ev) cannot be appreciably larger than the collision mean free path, l, of the heaviest components. We thus obtain from Table I:

$$L\lesssim4\times10^4 \text{ light years}.$$

The lower limit to L comes from considerations of intensity and isotropy. The density of cosmic-ray energy near the earth is of the same order of magnitude as the energy density of starlight. A mean life of cosmic rays in the galaxy approximately equal to that of starlight would then imply a cosmic-ray source of about the same strength as that of starlight. In order to avoid this very unlikely conclusion we must thus assume that $L\gg h$.

The argument based on the observed symmetry was first given by Cocconi,[5] and is here presented in a somewhat different form. One can estimate a lower limit for the anisotropy of cosmic rays at the earth by taking into account only the fact that the solar system is off the median plane of the galaxy by a distance $z_s=30$ l.y. (light years). For this purpose, consider the galaxy as a disk of thickness $2h$ and infinite extension, and assume that cosmic-ray sources are distributed symmetrically with respect to the median plane. Consider two planes parallel to the median plane, at distances z_s and $-z_s$ from it. Let Ψ_1 and Ψ_2 be the numbers of particles per unit area and unit time which cross one of the two planes entering or leaving the intervening volume, respectively. Thus the net number of particles per unit area and unit time leaving this volume through *both* boundaries is $2(\Psi_2-\Psi_1)$. If we neglect nuclear collisions, we then obtain the following equation expressing the conservation of particles:

$$2(\Psi_2-\Psi_1)=\int_{-z_s}^{+z_s}S(z)dz, \qquad (1)$$

[5] G. Cocconi, Phys. Rev. 83, 1193 (1951).

where $S(z)$ is the source strength in particles per unit volume and unit time. On the other hand, for a nearly isotropic distribution the following equation holds:

$$\Psi_1+\Psi_2=\tfrac{1}{2}\rho v, \tag{2}$$

where ρ is the density of particles and v their velocity. The ratio between the total number of particles present in the galaxy at any particular time and the number of particles produced per unit time represents the mean life, L/v, of the particles in the galaxy; thus

$$L/v=\int_{-h}^{+h}\rho dz \Big/ \int_{-h}^{+h} S(z)dz. \tag{3}$$

If we assume that the density of cosmic-ray particles near the earth is a fair sample of the mean density throughout the galaxy, the above equation becomes

$$L/v\approx 2h\rho \Big/ \int_{-h}^{+h} S(z)dz. \tag{4}$$

The asymmetry of cosmic rays with respect to the plane $z=z_s$ may be defined as

$$\delta=2(\Psi_2-\Psi_1)/(\Psi_2+\Psi_1). \tag{5}$$

From Eqs. (1), (2), and (4) one then finds for δ the expression

$$\delta\approx 4\frac{h}{L}\int_{-z_s}^{+z_s} S(z)dz \Big/ \int_{-h}^{+h} S(z)dz. \tag{6}$$

For example, in the case of a uniform distribution of sources through the galaxy, Eq. (6) reduces to

$$\delta\approx 4z_s/L. \tag{7}$$

At the mean energy of cosmic rays, about 10 Bev, $\delta\lesssim 10^{-3}$, and therefore

$$L\gtrsim 4000z_s=1.2\times 10^5 \text{ l.y.}$$

Even at energies as high as 10^{13} or 10^{14} ev, $\delta\lesssim 2\times 10^{-2}$, and $L\gtrsim 6000$ light years. These limits would be higher except for the fact that the sun lies close to the galactic equator. It is not likely that the local distribution of the sources is really as symmetrical as the over-all galactic disk model which we use, so that the limits given above are probably conservative.

It may be pertinent to point out that these general conclusions are valid irrespective of the specific mechanism opposing the escape of cosmic-ray particles from the galaxy. They apply equally well to models describing the galaxy as a volume filled with a diffusing medium, or as a volume bounded by a "white" wall. However, if one wanted to determine exact numerical values for the limits of L, rather than orders of magnitude, one would have to take into account the actual distribution of the sources, as well as the detailed mechanism of propagation of cosmic rays in the galaxy.

C. Energetics of the Cosmic Radiation

If we consider a radiation moving in the galaxy, either in straight paths like the starlight or in the tortuous paths of the diffusing cosmic-ray particles, we can write the following relation from energy conservation:

(energy density) = (source power density) × (mean life) × $(\bar{U}/\bar{U}_0)$,

where $\bar{U}_0$ is the mean energy per particle at emission and $\bar{U}$ the mean energy per particle observed in space. For starlight, $\bar{U}/\bar{U}_0=1$. Now it is plain that the mean life of starlight in the disk will be a few times h/c, and we have seen that for the bulk of the cosmic rays the mean life is L/c. It is known from astronomical data that in our location the energy density of light is approximately equal to the energy density of cosmic rays (both being about 1 ev per cm³). It therefore follows that the energy outputs of the cosmic-ray source and of the light source are in the ratio:

$$\text{cosmic rays/light}\approx h\bar{U}_0/L\bar{U}\approx 10^{-3}\bar{U}_0/\bar{U}.$$

Thus the injection of cosmic rays by stars would be only a small fraction of their total power output, even if there were no considerable acceleration in space. For our model, where the acceleration in space is decisive, we will see that $\bar{U}_0/\bar{U}\approx 10^{-3}$. The energy expenditure for the injection of cosmic rays is then an entirely negligible fraction of the stellar energy output. Even rare stellar events, such as supernovae, might supply the entire cosmic-ray beam injected.

On the other hand, our model does derive the energy of the cosmic-ray beam from interactions with the ion clouds in space, through their magnetic fields. The turbulent energy density of interstellar matter (both kinetic and magnetic) can be estimated roughly, and turns out to be of the same order of magnitude as the density of cosmic-ray energy, perhaps 1 to 10 ev per cm³. The energy presently stored in the interstellar turbulent motion could then supply the cosmic-ray power requirements only for a few million years, a time very short compared to the age of the galaxy. This implies that cosmic rays represent an important load on the energy input to the turbulent motion, which presumably is the galactic energy of rotation. In our neighborhood the ordered rotational energy density is about a thousand times greater than the turbulent energy.

Such considerations mean that the cosmic rays play an essential role in the dynamical evolution of a spiral galaxy. The energy loss to cosmic rays is an important factor in smoothing out the turbulence. We do not intend to discuss such problems here. We only wish to remark that according to the present picture, cosmic rays derive their energy from some major source of galactic nonnuclear energy, such as the gravitational energy gained by concentration toward the galactic

plane and toward the denser galactic center or, likely, some initial kinetic energy of rotation. Cosmic rays are presumably long-lasting features of a spiral galaxy which are not constant, but are related in time to the presence of well-marked, gas-laden, rotating arms.

II. THE DIFFUSION MODEL

A. The Diffusion Equation

We now wish to discuss a more specific model, by considering the galactic volume filled with wandering masses of turbulent, ionized, and magnetic hydrogen plasma moving in an un-ionized substratum. On this model, a cosmic-ray particle travels in nearly straight paths between magnetic clouds of streaming gas; upon entering such a cloud, the particle begins a tortuous trajectory, until it finally drifts out with almost no recollection of its original direction. To describe the random propagation of cosmic rays through the galaxy, we shall use a simple form of the transport theory, the diffusion theory, familiar in the study of slow neutrons. The low accuracy required does not suggest the use of any more elaborate method for the problem. It is worthwhile to remark that our values of such properties of the diffusing medium as the mean free path might well have to be modified if the distribution of types of scattering centers is an unusual one. We can regard the values given below as *effective* values of mean free path, etc., bearing a relation to such physical features as the mean spacing of scattering centers which is not exactly known. But for any reasonable collection of scattering objects, the broad features of diffusion theory still hold.

Fermi assumed that the scattering clouds have random motion and showed that in this case the particle energy will, on the average, increase during a collision. He also showed that the fractional energy gain by a particle in such a collision is a quantity $\alpha=\Delta U/U$ which depends on the velocity of the clouds, but not on U. Here, U is the total energy of the particle (kinetic+rest energy) expressed in units of its rest energy, m_0c^2. In most of our discussion we shall maintain the assumption that α is a constant, without thereby implying that the acceleration mechanism is identical to that originally suggested by Fermi.

We will denote by $\beta(U)$ the actual energy loss per collision (not the fractional loss), measured in units of the rest energy of the particle. This loss may arise from atomic collision in the plasma or in un-ionized gas, or from radiation accompanying accelerations in the magnetic field. We introduce, also, the number of collisions since the ejection of the particle into space, n. We regard the process—which involves n of some millions—as continuous, just as in the Fermi age theory of neutron slowing-down. The Fermi relation between energy and number of collisions is

$$dU/dn=\alpha U-\beta(U). \tag{8}$$

Evidently, particles injected at energies below a minimum energy U_e will not be accelerated:

$$U_e=\beta(U_e)/\alpha. \tag{9}$$

There is strong evidence that beams of cosmic-ray particles arrive at the earth during some solar flares. These particles have energies up to a few Bev per nucleon.[6] They come more or less directly from the neighborhood of the sun. It seems plausible to suppose that the sun is a continuous source of such particles, though perhaps with widely fluctuating intensity and energy spectrum. We may assume that most other stars, probably with still more extreme variations in number and energy of emission, also send fast particles into space (see Sec. IV C). Averaging over the stars, then, we will assign to the galaxy a distributed source density, capable of injecting into space protons, α particles, and other nuclei with energies far above the few kilovolts of the auroral streams, though still low compared to cosmic rays, or perhaps reaching at most a few Bev per nucleon. These emitted particles diffuse throughout the galactic disk, gaining and losing energy until they finally escape from the region of the plasma clouds, or rarely until they change their nuclear identity by a nuclear collision with the protons of space.

Let us write ρ for the number of particles per unit volume, at position $\mathbf{r}$ and time t, which have undergone between n and $n+dn$ collisions since their ejection from a source with energy U_0. Then the familiar scalar flux of particles per cm^2 and per second, useful in all nearly isotropic transport problems, is just $\varphi=\rho(\mathbf{r},t,n,U_0)v(n,U_0)$, where v is the velocity of the particles. The so-called "differential energy spectrum" of primary cosmic rays, j, gives the number of cosmic-ray particles per unit area, per unit solid angle, per second, and per unit energy interval. It is related to the flux φ by the equation

$$j(E)=(1/4\pi)\varphi dn/dE. \tag{10}$$

$E=U-1$ is the kinetic energy of a particle measured in units of its rest energy.

Following the familiar methods[7] of neutron diffusion and age theory, we obtain the differential equation expressing the conservation of particles:

$$\frac{\partial\rho}{\partial t}=\frac{\lambda}{3}\nabla^2\varphi-\frac{1}{\lambda}\frac{\partial\varphi}{\partial n}-\frac{\varphi}{l}+S(\mathbf{r},t,U_0)\delta(n). \tag{11}$$

Here λ is the transport mean free path for collisions with the clouds, and l the mean free path against nuclear collisions. We have written the source density function $S(\mathbf{r},t,U_0)$ for the number of particles injected per second per unit volume with initial energy between

[6] See the very valuable review by L. Biermann, Ann. Rev. Nuclear Sci. 2, 335 (1953).

[7] See, for example, S. Glasstone and M. Edlund, *The Elements of Nuclear Reactor Theory* (D. Van Nostrand Company, Inc., New York, 1952), Chap. VI.

U_0 and U_0+dU_0. The Dirac delta function indicates that the injection is, of course, with $n=0$. For each U_0, then, Eq. (8) gives a connection between U and n. By adding up the injections at all energies, one gets the complete spectrum. We shall generally neglect the initial spread of the injection spectrum.

B. Solution of the Diffusion Equation

It is enough to treat the stationary case, with $\partial\rho/\partial t=0$, and S constant in time. For any finite n, we have

$$\frac{\lambda}{3}\nabla^2\varphi-\frac{1}{\lambda}\frac{\partial\varphi}{\partial n}-\frac{\varphi}{l}=0. \tag{12}$$

Considering Eq. (11) in the neighborhood of $n=0$, we have

$$\varphi(\mathbf{r},t,0,U_0)=\lambda S(\mathbf{r},t,U_0). \tag{13}$$

No other boundary condition in the variable n is needed except the finiteness of φ for all n. The boundary condition to be set in ordinary space is less formal. It depends upon the reflectivity of the boundaries of the diffusing region. We shall set $\varphi=0$ at the extrapolated boundaries,[7] i.e., we shall assume no reflection. This assumption will be discussed later in detail (see Sec. IV B). Here we only remark that a moderate amount of reflection would not substantially alter the present treatment. Only a practically total reflection from the galactic boundaries would invalidate our conclusions.

For our problem it is convenient to introduce cylindrical coordinates with the origin at the center of the galactic disk and the equatorial plane coincident with the median plane. We denote the distance above the equatorial plane by z and the distance from the polar axis by R. We assume axial symmetry and symmetry about the equatorial plane as well.

One can solve Eq. (12) by the familiar method of series expansion into normal modes. For our geometry, this corresponds to an expansion into a Fourier series with respect to z and into a series of Bessel functions with respect to R. It turns out that the solution, or more specifically the functional dependence of φ on n at the location of the solar system, is not affected to any appreciable degree by the actual value of the galactic radius, R_0. This is easily understandable on physical grounds because $h\ll R_0$ and therefore the leakage through the cylindrical surface at $R=R_0$ is negligible compared with the leakage through the plane surfaces at $z=\pm h$. For this reason we can ignore the boundary condition $\varphi=0$ at $R=R_0$, i.e., we can consider the galaxy as having infinite lateral extension.

As a special and particularly simple case, we assume that the cosmic-ray sources are distributed uniformly in the radial direction, and symmetrically with respect to the equatorial plane. The problem then becomes unidimensional and the general solution of Eq. (12) obeying the boundary conditions $\varphi=0$ at $z=\pm h$ can be written as a Fourier series:

$$\varphi(z,n)=\sum_{i=1}^{\infty}A_i\cos[(i-\tfrac{1}{2})\pi z/h]\exp(-n/N_i), \tag{14}$$

where N_i is given by the equation

$$\frac{1}{N_i}=\frac{\pi^2(i-\frac{1}{2})^2\lambda^2}{3h^2}+\frac{\lambda}{l}. \tag{15}$$

In particular,

$$\frac{1}{N_1}=\frac{\pi^2\lambda^2}{12h^2}+\frac{\lambda}{l}. \tag{16}$$

The coefficients A_i are determined by Eq. (13) which, in this case, becomes

$$\lambda S(z)=\sum_{i=1}^{\infty}A_i\cos[(i-\tfrac{1}{2})\pi z/h]. \tag{17}$$

Thus the coefficients A_i are the Fourier coefficients of the source function multiplied by λ.

It is easy to modify the solution to allow for a possible variation of λ with n. It is necessary only to replace λn and $\lambda^2 n$ where they occur in the exponential of Eq. (14) by the integrals $\int_0^n\lambda(n')dn'$ and $\int_0^n\lambda^2(n')dn'$, respectively. The second integral is almost the "age," τ, in Fermi's theory of slowing-down:

$$3\tau=\int_0^n\lambda^2(n')dn'.$$

If we define the integral flux $\Phi(\mathbf{r},n)$ as

$$\Phi(\mathbf{r},n)=\int_n^{\infty}\varphi(\mathbf{r},n')dn', \tag{18}$$

we obtain from Eq. (14):

$$\Phi(z,0)=\sum_{i=1}^{\infty}N_iA_i\cos[(i-\tfrac{1}{2})\pi z/h]. \tag{19}$$

Notice that $\Phi(z,n)$ is the flux of particles at z which have undergone more than n accelerating collisions with the magnetic clouds of the galaxy, and $\Phi(z,0)$ is the total flux of particles. From Eqs. (3), (19), and (17) we obtain the following expression for the mean path L of cosmic-ray particles in the galaxy:

$$L=\int_{-h}^{+h}\Phi(z,0)dz\Big/\int_{-h}^{+h}S(z)dz$$

$$=\lambda\sum_{i=1}^{\infty}A_iN_i(i-\tfrac{1}{2})^{-1}\Big/\sum_{i=1}^{\infty}A_i(i-\tfrac{1}{2})^{-1}. \tag{20}$$

Unless the higher modes are important (i.e., unless the source strength varies rapidly in the z direction) L is of the order of magnitude of λN_1 and is, therefore,

given by

$$\frac{1}{L}\approx\frac{1}{\lambda N_1}=\frac{\pi^2\lambda}{12h^2}+\frac{1}{l}. \tag{21}$$

As discussed in Sec. I B, the mass spectrum of cosmic rays requires that $L\ll l$. Thus L must be of the order of magnitude of $12h^2/\pi^2\lambda$. This means physically, of course, that the mean life of cosmic-ray particles in the galaxy is primarily determined by the leakage through the two planes of the galactic disk. From the upper limit of L (4×10^6 l.y.) one finds a corresponding lower limit of λ: $\lambda>\frac{1}{4}$ l.y. We have also seen that isotropy considerations set a lower limit of about 10^5 l.y. to L. The corresponding upper limit of λ is 10 l.y.

If only the fundamental mode is present, the exact evaluation of δ becomes straightforward. The net flux of outgoing particles through the plane $z=z_s$ may be computed from Eq. (1) together with the expression for the source function:

$$S(z)=S_0\cos(\pi z/2h),$$

or from the equation:

$$\Psi_2-\Psi_1=-\frac{\lambda}{3}\left(\frac{\partial\Phi}{\partial z}\right)_{z=z_s}, \tag{22}$$

together with the expression for Φ:

$$\Phi(z,0)=\lambda S_0N_1\cos(\pi z/2h).$$

In either way one obtains (neglecting nuclear absorption):

$$\Psi_2-\Psi_1=(2/\pi)hS_0\sin(\pi z_s/2h). \tag{23}$$

On the other hand, Eq. (2) gives

$$\Psi_1+\Psi_2=(6h^2/\pi^2\lambda)S_0\cos(\pi z_s/2h). \tag{24}$$

Therefore, from Eq. (5):

$$\delta=(2\pi\lambda/3h)\tan(\pi z_s/2h)\approx\pi^2\lambda z_s/3h^2. \tag{25}$$

It happens that Eq. (25) is equivalent to Eq. (7), giving an approximate, but more general expression for δ. It may be pointed out that Eq. (25) applies not only to the asymmetry of the total flux Φ, but to that of the differential flux φ as well. In the latter case it remains valid also when λ is a function of n.

It is interesting to observe that another quite different and perhaps more plausible geometry for the galactic diffusing material will give nearly the same results as the disk described in detail here. The alternative is the confinement of the diffusion within a single spiral arm of the galaxy. Such an arm has a height $2h$, of course, a width perhaps $3h$, or $4h$, and a length comparable to R_0. The curvature can be neglected and the problem treated as diffusion in a long rectangular prism. Different eigenfunctions are appropriate, but the general features of the problem are the same, and the fundamental length is nearly the same as that given by Eq. (21). The additional leakage through the vertical bounding planes is only a correction to the already arge leakage through the top and bottom as before. Whether the fine structure of the galactic magnetic field shows the gaps between spiral arms as markedly as does the gas itself is not known. Since the qualitative results will be quite the same, we shall discuss the somewhat more evident case of disk diffusion in detail, but we do not by any means wish to obscure the possibility that the single-arm geometry, or some intermediate case, may in fact be the real solution.

C. Location of the Sources

In the previous section we have assumed that the sources are distributed uniformly in the plane of the galaxy (although they are not necessarily uniform perpendicularly to this plane). In order to investigate whether or not the details of the source distribution are important we consider next a source function of the following type:

$$S(\mathbf{r})=S_0\cos(\pi z/2h)\exp(-R^2/2D^2), \tag{26}$$

i.e., we assume that the source has a gaussian distribution with a rms spread D in the R direction, and is distributed according to the fundamental Fourier mode in the z direction. By integral expansion into Bessel functions, or by direct substitution into the diffusion Eq. (12) one finds the solution corresponding to the source function (26) for a galaxy of infinite radial dimensions:

$$\varphi=\lambda S_0\cos\left(\frac{\pi z}{2h}\right)\frac{6D^2}{6D^2+4\lambda^2n}\times\exp\left[-\frac{3R^2}{6D^2+4\lambda^2n}-\frac{n}{N_1}\right]. \tag{27}$$

We now further assume that $D\ll R$, i.e., that the dimensions of the source are small compared with the distance from the point of observation to the center of the source. The integral of Eq. (27) with respect to n can be expressed, after a suitable transformation, in terms of the Hankel function of zeroth order, $H_0^{(1)}$, so that the total flux of particles at R and z may be written as

$$\Phi(R,z,0)=\text{const}\times\cos(\pi z/2h)iH_0^{(1)}(2ia). \tag{28}$$

In the above equation, which is exact in the limit $D/R\to0$,

$$a=\frac{\pi R}{4h}\left[1+\frac{12h^2}{\pi^2\lambda l}\right]^{\frac{1}{2}}\approx\frac{\pi R}{4h}. \tag{29}$$

Figure 1 shows a graph of $H_0^{(1)}$ *versus* a. For $a\gg1$ Eq. (28) becomes approximately

$$\Phi(R,z,0)\approx\text{const}\times\cos(\pi z/2h)\times(2h/\pi R)^{\frac{1}{2}}\exp(-\pi R/2h). \tag{30}$$

One should note that, since we are considering a galaxy of infinite lateral extension, the line $R=0$ merely represents the axis of symmetry of the assumed source, so that the above equations apply to a source located anywhere in the galaxy at a distance R from the solar system. It is also plain that the results obtained do not essentially depend upon the particular form of the source density here used for calculation, provided

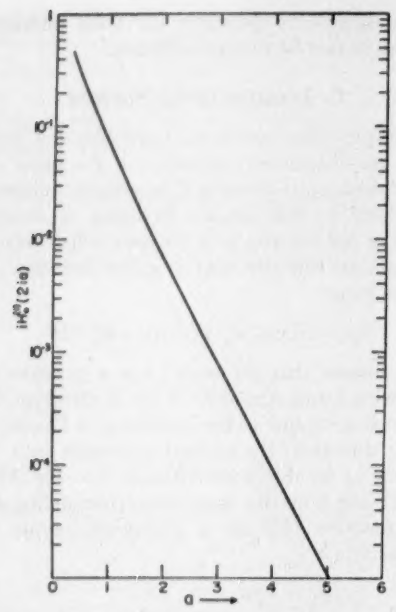

FIG. 1. The Hankel function, $iH_0^{(1)}(2ia)$, of zeroth order plotted *versus* $a=\pi R/4h$. $H_0^{(1)}$ describes the dependence of the total flux of cosmic-ray particles, $\Phi(R,z,0)$, on the radial distance, R, from the center of a concentrated source.

the radial dimensions of the source are small compared with R.

From Eq. (30) one sees that at distances larger than the thickness of the galaxy the contribution of a given source decreases rapidly with distance, the reduction factor being approximately $\exp(-\pi R/2h)$. It is thus difficult to escape the conclusion that the bulk of the cosmic rays—those with energies between 2 Bev and 10 Bev—are locally generated, in our galactic neighborhood, mostly within a few thousand light years from the sun. Any appreciable contribution to the total flux from sources further away than several times the thickness of the galactic disk would imply an extremely uneven distribution of the sources in the galaxy. It would also require an excessively large energy output from these sources. In particular, it seems difficult to assume, as it has been sometimes speculated, that most cosmic rays come from the dense concentration of stars at the center of the galaxy, 30 000 light years from the solar system. In our model the reduction factor for a source located at this distance relative to a local source is, approximately, 10^{-30}!

The above conclusions apply, as we have pointed out, to the bulk of the observed cosmic radiation. The high-energy end of the spectrum (large n) and the low-energy end of the spectrum (small n) require special consideration.

Equation (27) shows that φ does not decrease significantly with increasing R until R becomes larger than $\lambda\sqrt{n}$. If n is sufficiently large as to make $\lambda\sqrt{n}\gg h$, sources located at distances large compared with the thickness of the galactic disk may become as effective as local sources. One can easily see the physical reason for this result by considering that a particle arriving upon the earth after a large number of collisions has had approximately equivalent opportunities to escape from the galaxy whether it has begun its random walk near the earth or a considerable distance away.

To discuss the case of small n, consider again Eq. (27). If $\lambda\sqrt{n}$ is small compared with h, then φ decreases rapidly with increasing R even for distances small compared with the thickness of the galactic disk. Therefore, the observed flux depends critically on the details of the source distribution in the neighborhood of the sun.

One arrives at a similar conclusion considering the results of the preceding section relative to a source distribution depending only on z. For $n\gg N_1$, i.e., $n\gg h^2/\lambda^2$ [see Eq. (16)], the higher modes of the Fourier expansion of φ are negligible; i.e., φ has the same exponential dependence on n, $\exp(-n/N_1)$, irrespective of the source distribution. However, for $n\ll N_1$, the higher modes may be important and, therefore, the functional dependence of φ on n varies with the source distribution.

Note that the conclusions reached in this section are valid under any galactic theory of cosmic rays postulating a three-dimensional random-walk type of propagation through galactic space. In particular, distant sources, such as those situated at the center of the galaxy, could be important only if cosmic rays were confined to the galaxy for times long compared with h/c by a high reflectivity at the galactic boundary, rather than by frequent random changes in their direction of motion within the galaxy.

III. COMPARISON WITH EXPERIMENTS

A. The High-Energy Spectrum. Injection Energies

We have seen in the preceding section that for values of n larger than h^2/λ^2 (but not exceedingly large) the function φ is determined almost entirely by the average strength of cosmic-ray sources in the galactic neighborhood of the sun. In a theory which at best, can only account for the main features of the observed phenomena, we are therefore free to choose any reasonably smooth distribution of sources in the galactic disk. For ease of calculation, we shall assume that the source is distributed according to the fundamental mode in the z direction and is constant in the R direction:

$$S=S_0\cos(\pi z/2h). \tag{31}$$

In this case the solution of the diffusion equation is

$$\varphi=\lambda S_0\cos(\pi z/2h)\exp(-n/N_1), \tag{32}$$

and the integral flux is

$$\Phi(z,n)=\int_n^\infty \varphi(z,n')dn' =\lambda S_0 N_1\cos(\pi z/2h)\exp(-n/N_1). \tag{33}$$

The collision number n and the energy U are connected by Eq. (8). Let us first neglect the ionization loss and examine later whether or not this neglect is justified. We obtain

$$U=U_0e^{\alpha n};\quad n=\alpha^{-1}\ln(U/U_0). \tag{34}$$

From Eqs. (10), (33), and (34) we then compute the integral spectrum of cosmic rays, $J(E)$, i.e., the number of particles per unit area, per unit solid angle and per unit time, with kinetic energy greater than E:

$$J(E)=\int_E^\infty j(E')dE'=\frac{1}{4\pi}\Phi[n(E)]$$

$$=\frac{1}{4\pi}\lambda S_0N_1\cos\left(\frac{\pi z}{2h}\right)\left(\frac{U_0}{1+E}\right)^{1/\alpha N_1}. \tag{35}$$

This is the well-known power law of Fermi's theory.

Figure 2, taken from a paper by Barrett *et al.*,[8] summarizes our present information on the values of $J(E)$ at various energies from 14 Bev to about 2×10^6 Bev. The experimental data refer to all cosmic-ray particles. Little is known on the relative proportion of protons and heavier particles at the highest energies. It is, however, reasonable to assume that protons are greatly predominant here, as they are in the lower energy range. In what follows we shall discuss the data as if they applied entirely to protons. Note that the point at 14 Bev represents the vertical intensity at the geomagnetic equator, where the geomagnetic cutoff for protons is at 14 Bev kinetic energy. The other points result from observations of high-energy nuclear interactions in photoemulsions and from underground experiments on mesons. Data obtained from observations on air showers are also indicated (dashed lines). The measurements are not very accurate, but the range of energies is very great. Over this whole range the experimental data can be fitted to a power law of $(1+E)$ with exponent

$$-1/\alpha N_1=-(1.5\pm0.2). \tag{36}$$

The error refers to the uncertainty in the fit of the experimental data to an assumed power law; there is, of course, no assurance that this law is accurately verified.

Neglecting nuclear collisions, we can evaluate N_1 in terms of λ and h from Eq. (16):

$$N_1=12h^2/\pi^2\lambda^2\approx h^2/\lambda^2. \tag{37}$$

TABLE II. Critical kinetic energies of cosmic-ray particles.

Cosmic-ray particle:	Proton	α particle	CNO group	Fe group
Average atomic number	1	2	7	15
Critical kinetic energy (measured in Mev)	0.9	3.8	65	320

[8] Barrett, Bollinger, Cocconi, Eisenberg, and Greisen, Revs. Modern Phys. 24, 133 (1952).

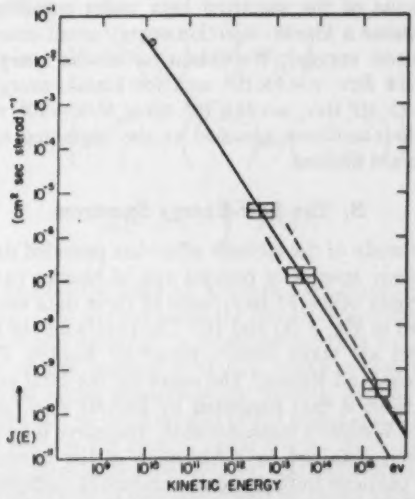

FIG. 2. The integral intensity of primary cosmic rays, $J(E)$, at kinetic energies from 14 Bev to 2×10^6 Bev. The experimental data indicated by squares are taken from a paper by Barrett *et al.* (see reference 8). The point at 14 Bev corresponds to the vertical integral intensity at the geomagnetic equator. The band between the two dashed lines indicates results deduced from the observations on air showers. The solid line represents the theoretical expression $J(E)=\text{const}\times(1+E)^{-1.5}$.

We then obtain from Eq. (36) the following relation:

$$\alpha=0.6\lambda^2/h^2. \tag{38}$$

If we consider the limits on λ discussed in Sec. II B and the fact that, on astronomical grounds, it is difficult to assign too large a value for α or too small a value for λ, we arrive at the following thoroughly tentative set of values

$$\lambda\approx1\text{ light year};\quad \alpha\approx6\times10^{-7}. \tag{39}$$

In Secs. IV A and IV D we discuss the plausibility of these assumptions. Now we must examine the consequences of the neglect of the ionization-loss term in the integration of Eq. (8). The actual ionization losses are rather uncertain. The loss rate in a completely ionized plasma is a good deal larger than that in un-ionized material essentially because of the greater value of the maximum impact parameter for effective collisions. But the division of the actual path traversed into plasma and un-ionized matter is not very well known. Most of the volume of the galactic disk is un-ionized. Taking for simplicity the loss rate in un-ionized hydrogen, we can compute the critical energies for injection of different nuclei from Eqs. (9) and (38). The results are given in Table II. Even if these energies are somewhat increased by the higher plasma losses, they are still very small compared to the mean cosmic-ray energies of 10 Bev or so. Thus the neglect of the ionization loss is justified.

From Eq. (34) and from the value of α given above we may compute the values of n corresponding to the

two limits of the spectrum here under consideration (we assume a kinetic injection energy small compared to the rest energy). We obtain for kinetic energies of about 14 Bev, $n \approx 4\times10^6$, and for kinetic energies of about 2×10^6 Bev, $n \approx 24\times10^6$. Since $\hbar^2/\lambda^2 \approx 10^6$, we see that the conditions specified at the beginning of this section are fulfilled.

B. The Low-Energy Spectrum

The study of the latitude effect has provided data on the energy spectra of protons and of heavier particles at energies below 10 Bev. Some of these data are summarized in Fig. 3 (a) and (b). The points for the heavy particles are taken from a paper by Kaplan, Peters, Reynolds, and Ritson.[9] The curve for the total number of particles is that suggested by Barrett *et al.*[8] on the basis of Winkler's measurements. The curve for protons has been obtained by subtraction of the numbers of heavy particles from the total numbers of particles. The dashed lines are drawn according to a power law of the type $J = \text{const}\times(1+E)^{-1.5}$.

It is necessary to emphasize the fact that most of the experimental results on the latitude effect are not very accurate. They are also difficult to interpret because the theory of geomagnetic phenomena is far from complete. However, it appears that while the data relating to the heavier components do not deviate significantly from the theoretical power law, the more accurate data relating to protons are in definite disagreement with it. Indeed, at energies around 0.1 Bev, the theoretical expression predicts about two or three times as many particles as are actually observed.

More recent results of Neher *et al.*[9] and of Van Allen[10] confirm this conclusion. Neher, working with ionization chambers at an atmospheric depth of 15 g cm^{-2}, finds an increase of only 1 percent in the cosmic-ray intensity between the geomagnetic latitudes of 56° to 66°. The corresponding kinetic cut-off energies for vertical protons are 0.8 Bev and 0.140 Bev, and according to the 1.5 power law, the increase should amount to approximately a factor of two. In a series of rocket experiments, Van Allen found that lowering the cut-off energy from 0.59 Bev to 0.018 Bev increased the intensity by only 14±9 percent as against a theoretical increase of about a factor of two. In both cases, the observed increases are hardly outside the experimental errors and, even if real, can be explained by the opening of the shadow cone as one proceeds toward higher latitudes. (Note that both Neher and Van Allen measured the omnidirectional intensity rather than the vertical intensity, which makes the interpretation of the results subject to greater uncertainty.)

In conclusion, there seems to be little doubt that, in the low-energy region, the proton spectrum becomes much flatter than the theory predicts. It is, however, impossible to say at the present time whether or not there exists an actual "cutoff," i.e., whether or not particles below a certain limiting energy (of the order of 0.8 Bev for protons) are totally absent from the primary spectrum. It is also uncertain whether or not the spectra of the heavier components exhibit a devia-

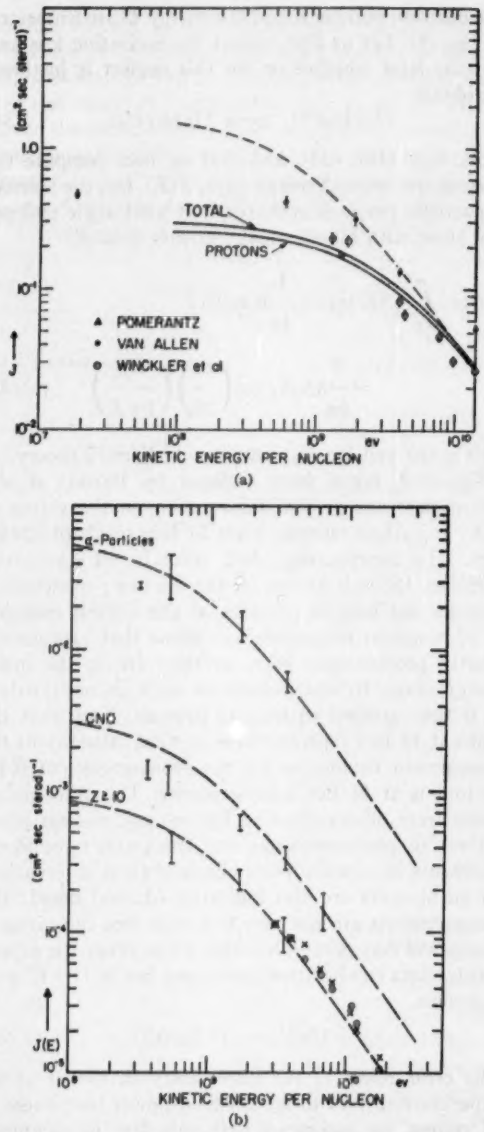

FIG. 3. The integral intensity of primary cosmic rays at energies less than 14 Bev per nucleon. (a) The upper solid curve is that suggested by Barrett *et al.* (see reference 8), and refers to all particles. The lower solid curve, deduced from the former by subtraction of the heavy particles, refers to protons. The circles and the triangles represent experimental determinations by Winkler and Pomeranz, as summarized by Puppi in Chapter VI of J. G. Wilson, *Progress of Cosmic Ray Physics* (North-Holland Publishing Company, Amsterdam, 1952). The dots represent measurements by Van Allen (see reference 10). The dashed line is drawn according to a power law of the type $J = \text{const}\times(1+E)^{-1.5}$ and is the continuation of the curve in Fig. 2. (b) The experimental points refer to measurements by Kaplon *et al.* (see reference 2), on the heavy components of the primary radiation. The dashed lines are drawn according to the power law $J = \text{const}\times(1+E)^{-1.5}$.

[9] Neher, Peterson, and Stern, Phys. Rev. **90**, 655 (1953).
[10] J. A. Van Allen, "The cosmic-ray intensity above the atmosphere near the geomagnetic pole," Department of Physics, State University of Iowa, January, 1953 (unpublished).

tion from the theoretical power law similar to that of the proton spectrum.

It has been suggested that the depletion of the low-energy end of the proton spectrum may not be a property of the primary beam but may be due to a shielding effect by local magnetic fields, such as a magnetic field of the sun.[11] The astronomical data, however, do not seem to support this view.[12] On the other hand, the simple diffusion model predicts, if anything, a deviation from the power law in the direction opposite to that observed. At low energies the ionization loss is not negligible and, in fact, the fractional energy gain per collisions will go to zero at the critical energy for injection, E_c. This effect would tend to *increase* the flux at low energies [see Eq. (10)]. Of course, in the absence of a detailed theory of the collisions between the particles and the magnetic elements, one cannot be sure that the simple equation $dU/dn=\alpha U-\beta$ with $\alpha=$const will hold down to the lowest energies.

One should also remark that, at energies slightly above the injection energy, the flux ought to be increased by particles injected above the critical energy and arriving without much acceleration. It is, of course, clear that if the source did not emit particles of energy less than $E_0 (E_0>E_c)$, no particles of energy less than E_0 would be found in the incoming radiation. Such an assumption, however, appears rather artificial.

In the absence of any other plausible interpretation of the low-energy end of the spectrum, it is perhaps interesting to note that the theory here under discussion can explain the observed features at the price, however, of introducing another *ad hoc* hypothesis. This hypothesis is that the source density is weaker than average in the neighborhood of the sun. If the sun is between two spiral arms, or within a gap even inside a main arm of the spiral, such an assumption would be reasonable. Formally, we could use as source density a function of the type

$$S=S_0\cos(\pi z/2h)[1-\epsilon\exp(-R^2/2\Delta^2)], \quad (40)$$

where we have chosen as central axis ($R=0$) a line passing through the sun and perpendicular to the galactic plane. The corresponding solution of the diffusion equation is [compare Eq. (27)]

$$\varphi=\lambda S_0\cos\left(\frac{\pi z}{2h}\right)\exp\left(-\frac{n}{N_1}\right)\times\left[1-\frac{\epsilon F}{F+n}\exp\left(-\frac{3}{4\lambda^2}\frac{R^2}{F+n}\right)\right], \quad (41)$$

where

$$F=3\Delta^2/2\lambda^2. \quad (42)$$

At the position of the solar system ($R=0$) the integral flux is

$$\Phi(z,0,n)=\lambda S_0 N_1\cos\left(\frac{\pi z}{2h}\right)\exp\left(-\frac{n}{N_1}\right)\times\left[1+\frac{\epsilon F}{N_1}\exp\left(\frac{F+n}{N_1}\right)Ei\left(-\frac{F+n}{N_1}\right)\right], \quad (43)$$

where

$$Ei(-x)=-\int_x^\infty e^{-t}dt/t$$

is the exponential integral. From this equation together with Eq. (34) one computes the integral spectrum. This can be easily fitted to the experimental data by proper choice of the parameter $(1-\epsilon)$, which measures the fractional source strength in our locality, and Δ, which gives the size of the local gap in the source distribution. If the sources are stars of some special type (like the T-Tauri stars suggested, wholly speculatively, in Sec. IV C) such a local gap might well be found astronomically. Only better experimental data will decide whether or not it is possible to fit the observations exactly with such a theory. It is at least satisfactory to see that Eq. (43) begins to give a deviation from the simple power law at an energy U_d such that $n(U_d)\approx F$. This implies [see Eqs. (34), (38), and (42)] that $\Delta\approx h[\ln(U_d/U_0)]^{\frac{1}{2}}$. Therefore, the size of the region of weak source would be of the order of h and vary quite slowly with U_d.

An alternative interpretation for the small number of low-energy particles is the assumption of a source of limited dimensions, situated some distance from the sun. But there is no plausible location for such a source apart perhaps from the galactic nucleus itself, which the intensity arguments of Sec. II C make very unsatisfactory. A local weakening of the source density is much more plausible than a relatively concentrated source somewhere a fraction of the galactic radius away from the sun.

C. Y. Fan,[13] in his account of the energy spectrum on Fermi's hypothesis derived the shape of the energy spectrum much as we have done here, though he neglected leakage. He ascribed the low-energy deviations to a finite distance from the source, just as we have done. He succeeded in fitting the whole spectrum by assuming a source in the dense Population II stars of the galactic nucleus and a *spherical* distribution of the diffusing magnetic field with dimensions of the order of R_0. He was required to postulate large injection energies and predicted for the heavy components an energy spectrum much steeper than for protons. We regard his results as sharply contradicted by cosmic-ray evidence alone, without reference to astrophysical plausibility.

C. Absence of Primary Electrons and Photons

It has been quite well established that the number of electrons and photons above 1 Bev arriving on the earth is less than 1 percent of the total particle flux.[14]

Electrons in the primary beam can have two origins: initial injection, together with the protons and the

[11] L. Janossy, Z. Physik **104**, 430 (1937).
[12] For a discussion of this question see also H. Alfvén, Arkiv Fysik **4**, 407 (1952).
[13] C. Y. Fan, Phys. Rev. **82**, 211 (1951).
[14] Critchfield, Ney, and Oleska, Phys. Rev. **85**, 461 (1952).

heavier components, or subsequent collisions of nucleons in space, producing mesons or neutrons. It has been estimated[15] that a few percent of the energy lost in collisions will appear in the form of electrons and photons resulting from the decay of mesons and neutrons. Since in our picture the total number of nuclear collisions suffered in space is quite small, we can expect from this source a negligible contribution to the electronic component. Indeed, the absence of electrons and photons re-enforces the nuclear mass distribution (the presence of the Fe group, doubtfully the absence of Li, Be, and B) in giving evidence that little matter has been traversed by the primary beam.

What of the possibility of electron injection? Our picture predicts that *total* energies are multiplied in the collisions with magnetic turbulence elements. If the electron spectrum at injection is like that of the heavy particles in the *kinetic energy scale*, then the final electron spectrum will be similar to the final proton spectrum, but with the energy scale reduced.

If the injection energy is, say, of the order of 10^7 ev, the reduction factor will be of the order of 100 and thus the electrons will be confined to very low energies. Such a similarity between the electron and the proton kinetic energies at injection would result from a single, more or less linear acceleration by a motional or induced emf. On the other hand, an injection mechanism which involves a cyclic or spiraling type of acceleration tends to produce electrons and protons with approximately the same momentum. In this case the reduction factor in the energy scale of the final spectrum is of the order of 10 and it may be necessary to postulate some mechanism to degrade the electron energy at a fast rate during propagation in space. Collisions with photons and radiation by deflection in magnetic fields have been suggested as possible processes leading to such energy losses.

In conclusion, unless electrons are in fact eliminated by some special mechanism, the theory suggests the existence of a primary electron component of intensity similar to that of the main beam, though of greatly reduced energy. A search for these electrons at high latitudes may be worthwhile.

D. Deviations from the Power Law at Extremely High Energies

For various reasons, one might expect deviations from the power-law energy spectrum at sufficiently high energies.

1. As the energy increases, the maximum distance at which sources are effective increases as well (see Sec. II C). When this distance becomes of the order of magnitude of the galactic radius, the assumption of an infinite uniform distribution of cosmic-ray sources in directions parallel to the galactic plane is obviously no longer justified. Indeed, there are no appreciable sources at $R>R_0$; moreover, near the center of the galaxy the strength of cosmic-ray sources may be far different from that in the neighborhood of the solar system. The maximum distance, R, at which the contribution of a source is important is related to the number of collisions, n, by the equation: $\lambda\sqrt{n}\approx R$. If we take for R half the galactic radius ($R=2.5\times10^4$ l.y.), we obtain for n a value of the order of 6×10^8. This corresponds to an energy of the order of $U_0e^{\alpha n}\approx U_0e^{360}$. We see that the cut-off energy set by the finite extent of the source is so high as to be practically meaningless.

2. Another conceivable cause for a cut-off is a time effect. Very high energy rays must have started a long time ago; on the present theory, the age of a particle is $\lambda n/v$, or about $10^6\times\ln(U/U_0)$ years. If the spiral arms evolve in times of the order of a few rotations, say 10^9 years, we could expect a cutoff for $\ln(U/U_0)\approx10^3$. Indeed, before the evolution of the spiral arms the rays may not have been contained in the galaxy to undergo acceleration. Here, too, however, the cut-off energy lies meaninglessly high.

3. The strict power-law for the energy spectrum follows from the assumption that the transport mean free path λ and the fractional energy gain per collision, α, are independent of energy. At a sufficiently high energy, fixed by the strength and size of the scattering magnetic elements, λ must increase and therefore the spectrum should fall off more rapidly, unless this effect is compensated by a large change in α.

Our experimental knowledge of the high-energy spectrum is not precise enough to exclude the possibility that λ may have already increased appreciably between 10^{10} and 10^{15} ev. In any case, as we shall discuss in Sec. IV A, the energy E_m, at which the increase of λ begins, could not lie much beyond 10^{15} ev. An increase of λ with energy should result not only in a departure from the power spectrum but also in a failure of isotropy.

Let us examine this point in some detail. Neglecting nuclear collisions, the differential energy spectrum for variable λ and α is given by the equation

$$j(E)=\frac{\text{const}}{E\alpha(E)}\cos\left(\frac{\pi z}{2h}\right)\exp\left[-\frac{\pi^2}{12h^2}\int_{E_0}^{E}\frac{\lambda^2(E')}{\alpha(E')}\frac{dE'}{E'}\right]. \quad (44)$$

On the other hand, the spatial asymmetry is given approximately by Eq. (25). One sees that the intensity at E depends on the integral

$$\int_{E_0}^{E}\lambda^2(E')dE'/\alpha(E')E',$$

while the spatial asymmetry is proportional to the value of λ at the given energy E.

Suppose first that α does not change much with energy. One will be able to observe a departure from isotropy before the intensity has dropped to an undetectable value if the increase of λ is sufficiently abrupt. To make a quantitative argument one should consider that the observations refer to integral rather than to differential intensities. For a power-law spectrum, the ratio, f, of integral intensities at the energies E and E_m is given by $f=Ej(E)/E_mj(E_m)$. For a more rapidly decreasing spectrum, such as corresponds to an increasing λ, f is smaller than the above

[15] See Phyllis Greifinger, Ph.D. thesis, Cornell University, Ithaca, New York, 1954 (unpublished).

expression. Therefore, from Eq. (44):

$$f<\exp\left[-\frac{\pi^2}{12h^2}\int_{E_m}^{E}\frac{\lambda^2(E')}{\alpha(E')}\frac{dE'}{E'}\right]=\exp\left[-\xi\frac{\lambda^2(E)}{h^2\alpha(E)}\left(\frac{E-E_m}{E}\right)\right], \quad (45)$$

where ξ is a number of the order of unity for any reasonable dependence of λ on E. Let us assume that λ is constant up to the energy $E_m=10^{15}$ ev. With air-shower detectors it may be possible to measure intensities 10^4 times smaller than that observed at 10^{15} ev. Setting, therefore, $f=10^{-4}$, the above equation yields

$$(E-E_m)/E<9\alpha h^2/\xi\lambda^2. \quad (46)$$

It does not seem likely that one can observe an anisotropy at these high energies unless λ is greater than $h/10$ (for this value Eq. (25) gives $\delta=0.01$; actually the anisotropy may be somewhat greater as already indicated in Sec. I B). One thus obtains

$$(E-E_m)/E<10^2\alpha/\xi.$$

Even if one takes into account the uncertainty in ξ, one sees that λ would have to increase by a factor of 100 in an energy range of less than 1 percent, which is very unlikely. One could of course expect an observable anisotropy if α at 10^{15} ev were much greater than at low energies, because then the required increase in λ would occur over a wider range of energies. Indeed, if we are willing to allow a large value of α at 10^{15} ev, we may assume that λ and α have been increasing smoothly with energy in such a way as to keep the ratio λ^2/α roughly constant. In this way we even may expect at some high energy a measurable anisotropy without appreciable departure from the power-law spectrum. Actually it would seem likely that α increases with energy (see Sec. IV A) but hardly fast enough to hold the intensity at a detectable level by the time a pronounced anisotropy is present. The above conclusions seem to be fairly independent of the details of the diffusion model. They suggest that the beam should remain essentially isotropic to the highest observable energies, if it is true (a) that the cosmic rays gain their energy through a gradual acceleration process, and (b) that the fractional energy gain per collision is not excessively large.

An experimental investigation of the angular distribution at the highest energies is obviously very desirable. Should anisotropy be found, it would require a drastic revision of the model. Moreover, it would furnish information on the geometry of the diffusing volume.

IV. ASTROPHYSICAL IMPLICATIONS OF THE MODEL

We shall now discuss in a perforce incomplete way the nature of the galactic magnetic field, and the astrophysical evidence bearing on the values of the parameters chosen to fit the cosmic-ray data.

A. The Transport Mean Free Path

One can construct without difficulty a model of a field in which, as required by our theory, the transport mean free path is constant up to a high limiting energy. As already mentioned, we picture the galactic space filled with wandering turbulent magnetic elements, each rather well separated from a background of lower field strength. The elements have a kind of hard core of higher gas densities and higher magnetic fields. Cosmic-ray particles wander from core to core in more or less straight-line paths; when they enter each core, they penetrate with a complex spiraling motion about the lines of force until they reflect from the region of strong field, and find their way out again. The mean free path will then depend primarily on the spacing of the cores, but *not* on the energy of the particle. This will remain true until the energy becomes so high that the radius of curvature, R_B, is comparable to the dimensions of the core. At this energy, E_m, the mean free path will increase and the power law will fail.

If one assumes spacings of the order of one or several light years and cores of somewhat smaller dimensions, E_m corresponds to a radius of curvature, R_B, of the order of 1 l.y. One can make a rough estimate of the magnetic fields in clouds by assuming equipartition of energy between magnetic fields and turbulent motion. If one takes one atom per cm^3 in the cores, and velocities of the order of 10 km/sec, one obtains for the magnetic field a value of the order of 3×10^{-6} gauss.[16] The corresponding value of E_m is about 10^{15} ev.

The picture of turbulence suggests that there are cores of different sizes and with different field strengths. As the energy increases a smaller and smaller fraction of the cores will remain able to scatter the particles. Thus, one should expect that beyond a certain energy, the mean free path will increase gradually, in a manner dependent on the distribution of the relevant properties of the cores.

From the crude argument of equipartition, one may expect that the cores of higher field strength will have higher average velocities. In Fermi's model, α is proportional to the square of the velocity. Therefore, one may predict an increase with energy of the fractional energy gain per collision, α, because particles of higher energies scatter preferentially from cores of higher field strengths and therefore higher velocities.

B. Reflectivity of Galactic Boundary

Reasons have been given[6,17] to assign a very large reflectivity to the boundary between the material of the galactic disk and the nearly field-free region outside. A reflectivity as great as 0.9 or 0.95 would have no important effect on our analysis but would require only a re-evaluation of such parameters as λ, h, etc. Only a reflectivity so great that a thousand or more approaches to the boundary should be made before escape, would drastically alter this model. It seems rather unlikely that such a high reflectivity can be maintained since (1) the ionized material is very spottily distributed, (2) it occupies a small fraction of the total volume of the disk or arm, and (3) it may indeed possess many lines of force which do not close, but point outward into extra-galactic space.

Another possible way of trapping cosmic-ray particles should be considered. The galaxy as a whole may possess a general external magnetic field, extending with appreciable intensity to a distance of the order of R_0 in all directions. This field, which may roughly resemble that of a dipole, will bend the trajectories of

[16] Neither the equipartition principle nor the values of density and velocity are reliable; but independent lines of evidence indicate galactic fields of this order of magnitude. See S. Chandrasekhar and E. Fermi, Astrophys. J. 118, 116 (1953).

[17] A. Unsöld, Phys. Rev. 82, 857 (1951).

the escaping particles and may bring them back to the galaxy after they have reached a distance of the order of R_0. In the slowly varying field the particles describe complicated trajectories, so that the time of drift T is long compared with the time of flight along the direct path T_0. Alfvén[18] has shown that T and T_0 are related by the approximate equation

$$T_0/T \approx R_B \partial(\ln B)/\partial r,$$

where B is the absolute value of the magnetic field, R_B is the radius of curvature, and $\partial B/\partial r$ is the derivative of B in the direction perpendicular to the lines of force. In a dipole field, $\partial(\ln B)/\partial r$ is of the order of the inverse distance from the center of the dipole and thus in our case of the order of $1/R_0$. If the particles are to be held in the neighborhood of the galaxy, R_B/R_0 must be small compared with one, say, of the order of $\frac{1}{10}$ at most. Since T_0 is at least R_0/c, the above equation shows that T is greater than 10^8 years. But, unless the particles return to cross the galactic disk some 1000 times in their life history, they will not greatly modify the characteristics of the diffusion process that we have postulated to account for the observed isotropy and intensity. This means that returning particles would have an average age in excess of 10^9 years, which is of the order of the age of the galaxy. Therefore, the properties of the galaxy in times far past are one of the factors that will determine the importance of an external trapping field on cosmic rays now present.

C. Mechanisms of Injection

The strongly reduced energies of injections of the present theory make simpler the problem of the mechanism of injection.

Acceleration of a fair sample of stellar atmosphere to energies of the order 100 Mev per nucleon seems a much more likely process than the production of particles with a mean energy of 10 Bev, and a tail up to 10^{13} ev. Indeed, the ejection of protons up to a few Bev during periods of great solar activity is now fairly certain, and it seems very reasonable to expect a much more constant and widespread mechanism giving rise to protons of considerably smaller kinetic energy even in stars of the type of the sun. The effect of solar flares also suggests that such stars as the T-Tauri variable, which show continually irregular flare-like variations, may be especially effective sources of cosmic rays.

Any electromagnetic mechanism for injection must avoid the heavy short-circuiting produced by an ionized plasma of considerable density.[19] This argument suggests that a sunspot betatron mechanism, such as suggested originally by W. F. G. Swann and elaborated by Butler and Riddiford[20] is not a very likely model if it is located deep in the stellar atmosphere.

We add a purely nuclear argument. Since the primary beam contains heavy nuclei, the paths of acceleration cannot have traversed much matter. The sunspot process involves times of the order of 10^3 or 10^4 seconds and therefore distances of some 10^{14} cm. Since the matter traversed must be less than a mean free path against nuclear collisions for a heavy nucleus, we require a density less than 10^{-13} g cm^{-3}. Such a density is not found below the outer chromosphere of the sun. Using the unjustifiable guide of energy equipartition, with densities of about 10^{-13} g cm^{-3} and velocities of streaming like those seen in sunspots (of the order of 100 km/sec), we get a maximum value for B of about 10 gauss. This seems to suggest that no mechanism which requires large fields for its action is likely to work with the matter densities required.

No dearth of other suggestions exists, the most likely of which use the large spaces and low densities available not in the stellar atmosphere proper, but in the region of the corona or beyond. In particular, the potentials induced by moving magnetized prominence-like jets of matter suggested by Kiepenhauer[21] and Schlüter[22] may be important sources of cosmic-ray particles. The high energies required without a rapid acceleration in space would place severe demands on the structure of the jets. On the other hand, the production of the comparatively low energies required by the present theory can be explained without great effort.

It is important to observe that the process considered here would also accelerate electrons. In the absence of strong magnetic fields it seems right to expect these electrons to enter the subsequent propagation as often as do protons. If there was no further acceleration in space, electrons and protons ought to arrive at the earth with comparable energies and intensities, unless some special process, such as those mentioned in Sec. III C, removes electrons very effectively during their propagation in the galaxy. The same conclusion applies to the case where particles are accelerated in space, but the rate of energy gain is so small as to require injection kinetic energies of the order of the rest energy of protons.

The suggestion here made, that a small fraction of the stellar energy output—of the order of 10^{-6}—goes into protons in the nuclear reaction range, and that this process occurs in regions far away from the regions of thermonuclear reaction, may be fruitful in other problems as well. F. Hoyle[23] has already suggested that such atmospheric processes may produce nuclear reactions enough to explain the presence in stars of such thermally perishable isotopes as deuterium.

[18] H. Alfvén, Z. Physik 105, 319 (1937).

[19] However, W. F. G. Swann has recently pointed out that the short-circuiting effect of ionized gases might have been overestimated [Duke Conference on Cosmic Rays, December 1953 (unpublished)].

[20] W. F. G. Swann, Phys. Rev. 43, 217 (1933); S. T. Butler and L. Riddiford, Phil. Mag. 43, 447 (1952).

[21] K. O. Kiepenhauer, Phys. Rev. 79, 809 (1950).

[22] A. Schlüter, Z. Naturforsch. 7a, 136 (1953).

[23] In conversation with E. E. Salpeter.

D. The Speed of the Magnetic Turbulence Elements

Fermi's original theory predicts a value of α equal to about $4V^2/c^2$, where V is the rms random velocity of the magnetic clouds. The high rate of acceleration required by our model would then indicate rms velocities of about 120 km/sec. Such velocities are beyond the maximum of some 80 km/sec, seen by Adams;[24] they are very much beyond his rms value of 30 km/sec and the even lower values, around 10 or 20 km/sec, favored in the later work. Moreover, the assumed small value of the transport mean free path (1 l.y.) goes against the general tendency to see rather larger clouds and a somewhat more uniform distribution of matter and of motion than this picture implies.

Without minimizing this difficulty we wish to submit the following remarks: All the astronomical studies are based on absorption or emission processes in space. The microwave emission studies, in particular, detect only unionized clouds, and do not directly reveal the motion of the plasma. Studies of interstellar spectral absorption lines are necessarily biased in favor of (1) clouds at high densities and large diameters which give stronger absorption, and (2) homogeneous motions, which lead to narrow and well-marked lines. Studies based on the residual polarization of starlight require dust, which probably accompanies rather dense accumulations of gas. Our clouds correspond to a relatively dilute swirling gas in which the denser and dusty clouds seen by the spectroscopists are drifting. The nature of turbulent motion in the plasma is not a problem beyond controversy, but a naive application of the Kolmogoroff spectrum of the homogeneous theory does imply that the more dilute clouds may have higher mean velocities. There may be some astrophysical evidence for this view.[25]

Recently Chandrasekhar and Fermi[26] have presented a picture of the galactic magnetic field as being fairly smoothly oriented along the spiral arms. Their picture is based on the supposed dust-grain orientation derived from polarization measures. Fermi[27] has shown that such a uniform field allows much more rapid acceleration than did his earlier picture of random collisions between particles and magnetic clouds. For the random case, it is only the excess of the head-on over the overtaking collisions which gives a net acceleration. But if the field is relatively smooth, a particle may be trapped along a line of force between two approaching clouds. These form "walls" between which (under certain conditions) repeated reflections will accelerate the particle until it gains enough energy so that it can break out of the trap. The velocities of the clouds do not determine the energy gain, but only the time required. This time can be a very small fraction of the history of the particle.

Fermi's new model accounts for the high rate of energy gain required by the cosmic-ray evidence, without postulating unobserved high velocities of the magnetic clouds. However, this model may be hard to reconcile with the high degree of isotropy of cosmic rays. Perhaps Fermi's acceleration mechanism, which makes use of an over-all regular magnetic field, is superposed on space diffusion through locally turbulent fields. The meaning of our parameter α will change in this more detailed picture: it will cover an average of acceleration rates both over the random collisions with the small, turbulent magnetic clouds which govern the motion in space and over the few but very effective collisions in the regions of uniform field, which govern the change in energy. But the theory as developed here would not be affected in its formulation. For it is from cosmic-ray properties and the general geometry of the gas distribution in the galactic disk alone that our conclusions follow.

[24] W. S. Adams, Astrophys. J. **109**, 354 (1949).
[25] S. B. Pikelner, Doklady Akad. Nauk S.S.S.R. **88**, 229 (1953).
[26] S. Chandrasekhar and E. Fermi, Astrophys. J. **118**, 113 (1953).
[27] E. Fermi, lecture at the meeting of the American Astronomical Society at Boulder, Colorado, August, 1953 (unpublished).

ACKNOWLEDGMENTS

We are grateful to Professor D. Menzel and to Drs. R. Thomas and M. Krook, all of the Harvard College Observatory, for their essential contributions to our astrophysical knowledge. They are responsible, however, for none of our errors. One of us (P.M.) is pleased to recall the kindness of Professors N. Frank and V. F. Weisskopf who made possible an enjoyable visit as Research Professor in their Department, during which this work was largely carried forward. We want also to thank the Cornell cosmic-ray group, Professor E. E. Salpeter of Cornell, and Professor B. Peters of Tata Institute, Bombay, for many stimulating discussions, both on cosmic-ray and on astronomical matters.

A Connection between Pion Photoproduction and Scattering Phase Shifts*

MARC ROSS
Brookhaven National Laboratory, Upton, New York
(Received December 21, 1953)

It is found in a rough approximation that meson photoproduction matrix elements can be expressed in terms of meson-nucleon scattering phase shifts and some energy-independent parameters, for photons in the laboratory of 200 to 500 Mev. The parameters are fitted to experiment in the lower portion of this energy region. It is then found that the higher-energy photoproduction data is consistent only with a $p_{\frac{3}{2}}$ isotopic spin $\frac{3}{2}$, phase shift which passes through 90° at a scattering energy of about 180 Mev. It is indicated that the s-wave phase shifts remain small.

INTRODUCTION

A SEMIPHENOMENOLOGICAL connection between photoproduction of mesons and meson-nucleon scattering phase shifts is developed in the energy region E_γ(lab) 200–500 Mev. It is assumed that the photoproduction matrix element to a particular angular momentum isotopic spin state is the sum of a Born approximation term plus a term in which the photoproduced meson is scattered before leaving the nucleon. The Born approximation portion of the matrix element is calculated in the (relativistic) weak coupling limit of the symmetric pseudoscalar coupling theory. The coupling constant $G^2/4\pi=16$ is selected beforehand to be in agreement with the (unrenormalized) Tamm-Dancoff theory of meson scattering.[1] The other portion of the photoproduction matrix element will be expressed in terms of parameters which can be fitted to experiment in the low-energy region where fairly complete data on scattering and photoproduction are available.† At higher energies, where scattering experiments have not as yet led to unique phase shifts, the behavior of the phase shifts can be determined from photoproduction data.

PHOTOPRODUCTION MATRIX ELEMENTS

Let $l_{j,m}{}^{0,+}$ be the photoproduction matrix element to a state of orbital angular momentum l, total angular momentum j, and its projection m, with 0,+ being the type of meson produced from photons on protons, such that the differential cross section is

$$\frac{d\sigma^{0,+}}{d\Omega}=\sum_m\langle|\sum_{l,j}l_{j,m}{}^{0,+}\phi_{lj,m}(\theta,\sigma)|^2\rangle_\sigma \tag{1}$$

in microbarns/sterad in the center-of-mass system, where we define

$$\phi_{s\frac{1}{2},\frac{1}{2}}=\chi^+,$$

$$\phi_{p\frac{1}{2},\frac{1}{2}}=-\cos\theta\chi^+-\sin\theta e^{i\phi}\chi^-,$$

$$\phi_{p\frac{3}{2},\frac{1}{2}}=\sqrt{2}\cos\theta\chi^+-\frac{1}{\sqrt{2}}\sin\theta e^{i\phi}\chi^-,$$

$$\phi_{p\frac{3}{2},\frac{3}{2}}=\sqrt{\tfrac{3}{2}}\sin\theta e^{i\phi}\chi^+.$$

Higher angular momentum states ($l>1$) are treated in a group as a function of θ, m, σ. Thus,

$$\sum_{l>1,j}l_{j,m}{}^{0,+}\phi_{lj,m}=\sum_s K_{m,s}{}^{0,+}(\theta)\chi^s. \tag{2}$$

To obtain the matrix elements $l_{j,m}$ consider the wave function of the one-nucleon one-meson component of the final state to have the form

$$\sum_\Gamma\psi_\Gamma\underset{r\to\infty}{=}\sum_\Gamma e^{i\delta}i^l\frac{\sin(kr-\frac{1}{2}l\pi+\delta)}{kr}\phi_\Gamma, \tag{3}$$

with the ϕ_Γ a set of orthonormal angular momentum isotopic spin functions, where Γ represents the appropriate quantum numbers. δ is the scattering phase shift in the state Γ, k is the center-of-mass momentum. The explicit dependence of the photoproduction on Γ is given in Eq. (3).

It is easily seen that

$$\int\left(\frac{k'}{k}\right)^2dE'\frac{1}{E-E'}\frac{\sin(k'r-\frac{1}{2}l\pi)}{k'r}\underset{r\to\infty}{=}\pi\frac{\cos(kr-\frac{1}{2}l\pi)}{kr}, \tag{4}$$

where E is the energy of the system. Then we can write (for all r)

$$\psi_\Gamma=\int\left(\frac{k'}{k}\right)^2dE'e^{i\delta}\cos\delta\left[\delta_{(E-E')}-\frac{g_E(E')}{E-E'}\right]i^lf_l(k'r)\phi_\Gamma, \tag{5}$$

where, from Eqs. (3) and (4), $g_E(E)=-\pi^{-1}\tan\delta$. The function $g_E(E')$ for $E'\neq E$ determines the wave function at small distances. The quantity in the bracket is seen to be the Γth partial wave of the standing wave solution

* This work was performed under the auspices of the U. S. Atomic Energy Commission.

[1] Dyson *et al.* (to be published); F. J. Dyson, in *Proceedings of the Rochester Conference on High Energy Physics, December, 1952* (Interscience Publications, New York, 1953); Sundaresan, Salpeter, and Ross, Phys. Rev. 90, 372 (1953).

† *Note in proof.*—General properties of the problem which are independent of this model are discussed by K. Aizu, Proceedings of the International Conference on Theoretical Physics, Tokyo (Sept. 1953), and K. M. Watson, Phys. Rev. (to be published). A formal treatment of the model in question is being prepared by the author [for a brief discussion see M. Ross, Phys. Rev. 92, 855 (1953)]. For earlier phenomenological work see K. A. Brueckner and K. M. Watson, Phys. Rev. 86, 923 (1952).

of the Goldberger formalism.[2] When the matrix element is computed with ψ_V, the delta function term yields $e^{i\delta}\cos\delta B_E(E)$ while the other term is $e^{i\delta}\cos\delta \int dE'(k'/k)^2 \times B_E(E')g_E(E')/(E-E')$, where $B_E(E')$ is the Born approximation matrix element to the state Γ with off the energy shell components indicated by $E'\neq E$.

The photoproduction matrix element to a particular angular momentum isotopic spin state can then be written

$$(B_E(E)\cos\delta + A\sin\delta)e^{i\delta}, \tag{6}$$

where

$$A=\frac{1}{\tan\delta}\int\left(\frac{k'}{k}\right)^2 dE'\frac{g_E(E')}{E-E'}B_E(E').$$

In order to calculate these matrix elements restrictive assumptions must be made. As stated we shall calculate the Born approximation in lowest-order perturbation theory. In this theory B_E is roughly independent of the energy E.[3] We shall assume this independence to be exact. In addition, it is assumed that $g_E(E')/(E-E')$, the scattered wave in the meson scattering problem, has the same shape, or relative dependence on the variable $E-E'$, independent of E. Only the magnitude is considered to depend on energy. The magnitude of g is assumed to behave as $g_E(E)$,[4] i.e., as $\tan\delta$, so we have that A is roughly independent of E and δ. It should be kept in mind that this is a crude approximation.

The details of the calculation of photoproduction in the weak coupling limit can be found elsewhere.[5] We shall merely state results here. In the problem at hand we require the photoproduction matrix elements $B_E(E)$ and an analysis into their s and p components. The following notation is employed: $\hbar=c=1$. Results are stated in the center-of-mass system, with the Z axis given by the incoming photon. Let μ = meson mass, M = nucleon mass, $\mathbf{k}$ = meson momentum, θ = angle between $\mathbf{k}$ and Z, $\omega=(k^2+\mu^2)^{\frac{1}{2}}$, $\epsilon=(k^2+M^2)^{\frac{1}{2}}$, and $E=\omega+\epsilon$. The matrix elements to the plane-wave final state are functions of θ and are characterized by the projections of final spin and of total angular momentum, s and m; and by the isotopic spin state. The polarization of the initial photon is taken to be $\mathbf{x}+i\mathbf{y}$.

In the symmetric pseudoscalar coupling theory[6] the photoproduction proceeds in lowest order by the three diagrams of Fig. 1. The matrix elements associated with these diagrams are

$$B^{(1)}=CI^{(1)}\frac{\sqrt{2}a}{-\omega+k\cos\theta},$$
$$B^{(2)}=-CI^{(2)}\frac{a+b}{\epsilon+k\cos\theta}, \tag{7}$$
$$B^{(3)}=CI^{(3)}\frac{b}{\sqrt{2}E},$$

where

$$C=-\frac{ieG}{4\pi}\left(\frac{k(\epsilon+M)}{2(E^2-M^2)}\right)^{\frac{1}{2}}\frac{1}{E-M},$$
$$a=\frac{k\sin\theta e^{i\phi}}{\sqrt{2}}\left(\chi^s,\left[-\frac{\mathbf{k}\cdot\boldsymbol{\sigma}}{\epsilon+M}+\frac{E-M}{E+M}\sigma_z\right]\chi^{M-1}\right),$$
$$b=\frac{E-M}{2}\left(\chi^s,\left[\frac{\mathbf{k}\cdot\boldsymbol{\sigma}}{\epsilon+M}+1\right]\frac{\sigma_x+i\sigma_y}{\sqrt{2}}\chi^{M-1}\right),$$

and the I's denote the appropriate isotopic spin factors. If each I is written as the coefficient in the isotopic spin $\frac{1}{2}$ state and the coefficient in the $\frac{3}{2}$ state, respectively, then

$$I^{(1)}=(-\sqrt{\tfrac{2}{3}},\sqrt{\tfrac{1}{3}});\quad I^{(2)}=(\sqrt{\tfrac{1}{3}},\sqrt{\tfrac{2}{3}});\quad I^{(3)}=(\sqrt{6},\,0).$$

The matrix elements to final s and p states are[7]

$$s_{\frac{1}{2}}=\frac{C}{2}\left[-I^{(1)}\frac{k}{\epsilon+M}J-I^{(2)}\frac{\sqrt{2}}{\epsilon}\left(\omega+\frac{\epsilon-M}{3}\right)+I^{(3)}\frac{E-M}{E}\right],$$
$$p_{\frac{1}{2}}=\frac{C}{2}\left[I^{(1)}\frac{E-M}{E+M}J-I^{(2)}\frac{(E-M)k}{3\sqrt{2}\epsilon^2}\left(1-\frac{\omega}{\epsilon}\right)-I^{(3)}\frac{(E-M)k}{E(\epsilon+M)}\right],$$
$$p_{\frac{3}{2},\frac{1}{2}}=\frac{C}{2}\left[\frac{I^{(1)}}{\sqrt{2}}\left\{\left(\frac{E-M}{E+M}-\frac{3\omega}{\epsilon+M}\right)J-\frac{4k}{\epsilon+M}\right\}+I^{(2)}\frac{(E-M)k}{3\epsilon^2}\right], \tag{8}$$
$$p_{\frac{3}{2},\frac{3}{2}}=\frac{C}{2}\left[\sqrt{\tfrac{3}{2}}I^{(1)}\left\{\left(\frac{E-M}{E+M}-\frac{\omega}{\epsilon+M}\right)J-\frac{4k}{3(\epsilon+M)}\right\}+\frac{I^{(2)}}{2\sqrt{3}}\left(-\frac{4(E-M)}{E+M}-\frac{\epsilon-M}{\epsilon}\right)\frac{k}{\epsilon}\right],$$

[2] M. L. Goldberger, Phys. Rev. **84**, 929 (1951).

[3] See for example the total cross-section curve in G. Araki, Progr. Theoret. Phys. (Japan) **5**, 570 (1951), or the $d\sigma^+/d\Omega(90°)$ curve in J. Steinberger and A. S. Bishop, Phys. Rev. **86**, 171 (1952).

[4] Some indication that this may be reasonable comes from the the calculation of g in the Tamm-Dancoff approximation (reference 1). Where it is found in those cases examined, that for E near threshold, $g_E(E')$ is fairly flat from say $E'=E$ to $E+Mc^2$ (where M is the nucleon mass), while it drops sharply as E' decreases below E.

[5] Benoist-Gueutal, Prentki, and Ratier, Compt. rend. **230**, 1146 (1950); G. Araki, Progr. Theoret. Phys. (Japan) **5**, 507 (1950); K. A. Brueckner, Phys. Rev. **79**, 641 (1950); M. F. Kaplon, Phys. Rev. **83**, 712 (1951).

[6] We consider the interaction $iG\bar{\psi}\gamma_5\tau_\alpha\phi_\alpha\psi$. The coupling constant G is $\sqrt{2}$ times that usually employed in the charged theory. The latter is used by Kaplon (reference 5), and Steinberger and Bishop (reference 3), among others.

[7] The contribution from $B^{(3)}$ which is relatively small has been approximated for convenience. The first two terms in an expansion in powers of k/E are kept, with a consequent error of $\lesssim 5$ percent in the $B^{(3)}$ contribution in the energy region under consideration.

where

$$J=k\int_{-1}^{1}\frac{(1-x^2)dx}{-\omega+k\cos\theta}=2\left[\left(\frac{\mu}{k}\right)^2\ln\left(\frac{k+\omega}{\mu}\right)-\frac{\omega}{k}\right].$$

One can also calculate the wave function g, and thus the terms A in the photoproduction matrix elements in some approximation, from the meson theory. It is felt, however, that this is a reasonable place for a phenomenological viewpoint. The A's are to be regarded as parameters determined experimentally. Let a_1, a_3, c, d be proportional to the parameters A in the $s_{\frac{1}{2}}$, isotopic spin $T=\frac{1}{2}$; $s_{\frac{1}{2}}$, $T=\frac{3}{2}$; $p_{\frac{1}{2},\frac{1}{2}}$, $T=\frac{3}{2}$; $p_{\frac{1}{2},\frac{3}{2}}$, $T=\frac{3}{2}$ states, respectively. It is assumed for simplicity that the phase shifts in the other states are small enough so that the corresponding parameters in these states exert negligible influence.

If the Born approximation from Eq. (8) at $E_\gamma(\text{lab})=310$ Mev[8] is used, the photoproduction matrix elements in this theory are

$$\begin{aligned}
s_{\frac{1}{2}}{}^{+}&=e^{i\delta_1}(-1.27\cos\delta_1+a_1\sin\delta_1)\\
&\quad+e^{i\delta_3}[-0.94\cos\delta_3+(a_3/\sqrt{2})\sin\delta_3],\\
s_{\frac{1}{2}}{}^{0}&=e^{i\delta_1}(+0.90\cos\delta_1-(a_1/\sqrt{2})\sin\delta_1)\\
&\quad+e^{i\delta_3}(-1.32\cos\delta_3+a_3\sin\delta_3),\\
p_{\frac{1}{2}}{}^{+}&=-0.41-0.43=-0.84,\\
p_{\frac{1}{2}}{}^{0}&=+0.29-0.62=-0.33,\\
p_{\frac{3}{2},\frac{1}{2}}{}^{+}&=-0.34+e^{i\delta}(-0.06\cos\delta+(c/\sqrt{2})\sin\delta),\\
p_{\frac{3}{2},\frac{1}{2}}{}^{0}&=+0.24+e^{i\delta}(-0.09\cos\delta+c\sin\delta),\\
p_{\frac{3}{2},\frac{3}{2}}{}^{+}&=-0.70+e^{i\delta}(-0.52\cos\delta+(d/\sqrt{2})\sin\delta),\\
p_{\frac{3}{2},\frac{3}{2}}{}^{0}&=+0.50+e^{i\delta}(-0.73\cos\delta+d\sin\delta).
\end{aligned}\tag{9}$$

As in the $p_{\frac{1}{2}}$ state for both isotopic spins, in the $p_{\frac{3}{2}}$ state the $T=\frac{1}{2}$ phase shift is assumed to be sufficiently small to be neglected. The symbol δ indicates the $p_{\frac{3}{2}}$, $T=\frac{3}{2}$ phase shift. In the s state δ_1, δ_3 are for $T=\frac{1}{2}$, $\frac{3}{2}$, respectively.

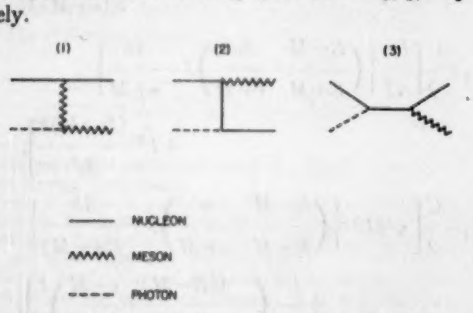

FIG. 1. Lowest order diagrams for photoproduction.

[8] Convenient formulas for the laboratory energies in which the initial proton is at rest, $E_\gamma(\text{lab})$, and in which the final nucleon is at rest, $E_\pi(\text{lab})$, are $E_\pi(\text{lab})=(\mu^2+k^2E^2)^{\frac{1}{2}}-\mu$, $E_\gamma(\text{lab})=E_\pi(\text{lab})+\mu[1+(\mu/2M)]$.

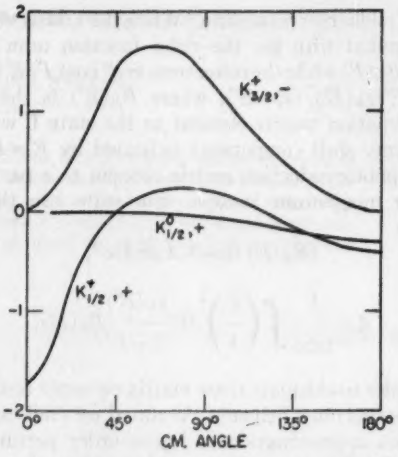

FIG. 2. High angular momentum contribution to Born approximation matrix elements.

Those higher angular momentum matrix elements which are not trivially small are deduced from Eqs. (7), (8) and are expressed in terms of $K_{m,s}{}^{0,+}(\theta)$, as in Eq. (2). They are plotted in Fig. 2.[9] The phase shifts in these states with $l>1$ are assumed to be sufficiently small so that the Born approximation is valid.[10]

We have some useful *a priori* knowledge of the parameters. Thus, an attractive phase shift will generally increase the Born approximation matrix element to a given angular momentum isotopic spin state. The only deviations from this rule, that might be expected to occur, would be associated with competition of processes. Care must be exercised in the two $p_{\frac{1}{2}}$ states where $E2$ and $M1$ interactions compete. It is found that a_1, a_3, d should be negative, and $c\doteq-d/\sqrt{3}$.[11]

The differential cross sections are calculated from Eq. (9) and Fig. 2, using Eq. (1). An example will be given here because of its particular importance. Thus,

$$\frac{d\sigma^0}{d\Omega}(90°)=\tfrac{2}{9}|p_{\frac{3}{2},\frac{3}{2}}|^2+\left|\frac{1}{\sqrt{2}}p_{\frac{3}{2},\frac{3}{2}}+p_{\frac{1}{2}}\right|^2+|s_{\frac{1}{2}}|^2.\tag{10}$$

In this special case a convenient approximation can be made. The $p_{\frac{3}{2},\frac{3}{2}}$ contribution is found to be about 90 percent of the r.h.s. of Eq. (10). Furthermore, it is easily shown that $p^0{}_{\frac{3}{2},\frac{3}{2}}\sim\sin(\delta+\alpha)$ within about 1 per-

[9] While neutral production can be expressed as $A+B\cos\theta+C\sin^2\theta$, it is seen that this will not be strictly true for positive mesons. At very small angles large deviations from this type of behavior can occur.

[10] Errors introduced in neglecting any small phase shifts are small in photoproduction, though a small d wave phase shift may loom large in the analysis of a scattering distribution.

[11] This relation holds between the matrix elements for a pure $M1$ transition, while the relation $p_{\frac{3}{2},\frac{1}{2}}=\sqrt{2}p_{\frac{3}{2},\frac{3}{2}}$ would hold for an $E2$ transition. With a strong attractive interaction one expects the $M1$ transition to be strongly enhanced over the $E2$. The weak coupling π^+ matrix element in the $p_{\frac{3}{2},\frac{1}{2}}$ state happens to correspond to the $E2$ interaction being stronger than $M1$, so that c has the opposite sign to the Born approximation.

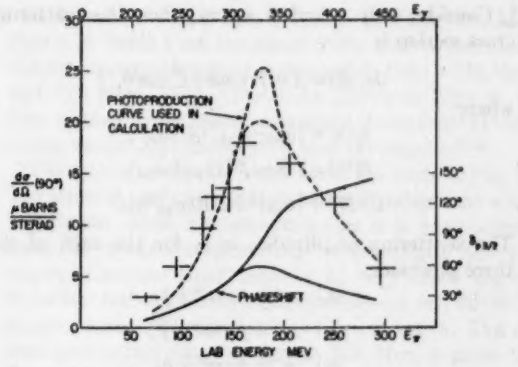

FIG. 3. Neutral photoproduction and the corresponding $p_{\frac{1}{2}}$, $T=\frac{1}{2}$ phase shift from Eq. (11). The experimental points (reference 13) are at 90° in the laboratory, while the theoretical points are in the center of mass.

cent. Then the simple relation,

$$\frac{d\sigma^0}{d\Omega}(90°)\sim\sin^2(\delta+\alpha), \tag{11}$$

where

$$\alpha=\tan^{-1}\left(\frac{0.50-0.73}{d}\right),$$

can be used to determine the $p_{\frac{1}{2}}$, $T=\frac{1}{2}$ phase shift with fair accuracy.

COMPARISON WITH EXPERIMENT

Using measured scattering phase shifts for E_π(lab) $\leq$135 Mev[12] and fitting $d\sigma^0/d\Omega(90°)$ at E_γ(lab)=285 Mev (Fig. 3),[13] $d\sigma^+/d\Omega(\theta)$ at 265 Mev (Fig. 4)[14–16] and $d\sigma^+/d\Omega(90°)$ *vs* E (Fig. 5) up to 265 Mev, we obtain the parameters

$$d=-4,\quad \alpha=3\tfrac{1}{2}°,\quad c=2,\quad a_3-a_1=1,$$

(with the s-wave parameters especially rough). The π^0 angular distribution could also be helpful in determining the parameters, unfortunately it is now known only roughly. In the neighborhood of 250–300 Mev it is seen that the quantity (approximately) $2a_3+a_1$ determines the asymmetry or $\cos\theta$ term in the π^0 angular distribution. If we assume arbitrarily that there are 10 percent more π^0 forward of 90° than backward,‡ then (roughly)

$$a_3=-1,\quad a_1=-2.$$

The $p_{\frac{1}{2}}$, $T=\frac{3}{2}$ phase shift predicted (above $E=135$ Mev) in accordance with Eq. (11) is shown in Fig. 3. There are two types of behavior. If the maximum in the neutral meson photoproduction at 90° is as high as *about* 25 millibarns/sterad (upper curve), then the $p_{\frac{1}{2}}$ phase shift can pass over 90° (case I). If the photoproduction maximum is lower, then the $p_{\frac{1}{2}}$ phase shift decreases before reaching 90° (case II, lower curve). The experimental values for $d\sigma^0/d\Omega(90°)$ are too rough, as yet, to determine the significant parameter d or to place the maximum with sufficient accuracy to distinguish between the cases.

It is seen that between E_π(lab)=150 Mev and 210 Mev, the detailed behavior of phase shift is not given by the measured values of $d\sigma^0/d\Omega(90°)$. In this region, *in order to be definite*, δ was chosen to pass through 90° at 180 Mev and cotδ was required to vary smoothly

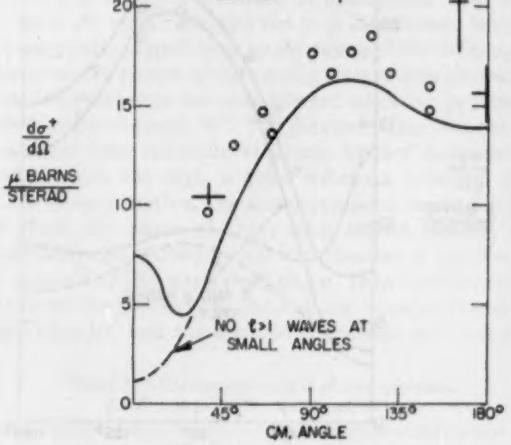

FIG. 4. Photoproduction of π^+ at 265 Mev.

[12] The scattering phase shifts used in the analysis are those of Anderson, Fermi, Martin, and Nagle, Phys. Rev. **91**, 155 (1953), and Bodansky, Sachs, and Steinberger, Phys. Rev. **90**, 997 (1953).

[13] A complete survey of the π^0 experimental situation was not made. The first two of the following references were arbitrarily chosen for comparison. A. Silverman and M. Stearns, Phys. Rev. **88**, 1225 (1952); Walker, Oakley, and Tollestrup, Phys. Rev. **89**, 1301 (1953); Goldschmidt-Clermont, Osborne, and Scott, Phys. Rev. **89**, 329 (1953); C. G. Andre, UCRL 2425 (1953).

[14] Preliminary results of P. D. Luckey and R. R. Wilson: π^+ photoproduction measurements up to 280 Mev.

[15] Preliminary results of Cal-Tech group (R. L. Walker *et al.*): π^+ photoproduction differential cross sections from 200 to 500 Mev. Some π^0 differential cross sections at high energies have also been measured by this group.

[16] A complete survey of the π^+ experimental situation was not made. References 14 and 15 were arbitrarily chosen for comparison. Experimental work on π^+ photoproduction has also been done by Steinberger and Bishop (reference 3); White, Jakobson, and Schulz, Phys. Rev. **88**, 836 (1952); Jarmie, Repp, and White, Phys. Rev. **91**, 1023 (1953); Goldschmidt-Clermont, Osborne, and Winston, Phys. Rev. **91**, 468 (1953); E. L. Goldwasser and G. Bernadini, Bull. Am. Phys. Soc., Vol. 29, No. 1, 18 (1954); G. Sargent Janes and W. L. Kraushaar, Phys. Rev. **93**, 900 (1954). With reference to the first three experiments above (from Berkeley), there has been some question of a difference in the calibration used relative to that used at other laboratories. It is believed that there is no such difference here (within 5%) (private communication with R. S. White).

‡ Preliminary indications are that this is incorrect (L. S. Osborne, private communication). The results that follow will not, however, be too sensitive to this behavior.

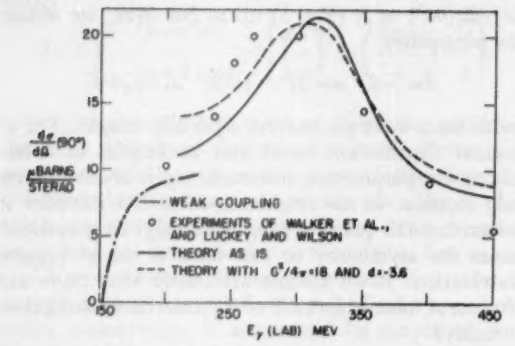

FIG. 5. Photoproduction of π^+ at 90° in the center-of-mass system. (See reference 20.)

with energy. What we shall call the resonance (the passing through 90° by the phase shift) could occur anywhere from say 165–195 Mev and could be considerably narrower or broader than that shown, without straining the fit to the existing π^0 data.

The most sensitive test of whether this resonance occurs (case I) or whether the phase shift remains below 90° (case II) is given by $d\sigma^+/d\Omega$ at higher energies. Since the π^- scattering data[17] at these energies antedates the photoproduction experiments we shall use the former to determine the s-wave phase shifts involved in the latter. The determination of the phase shifts from the scattering data is easy, and in its essentials quite unambiguous, now that the $p_{\frac{1}{2}}$, $T=\frac{3}{2}$ phase shift is given

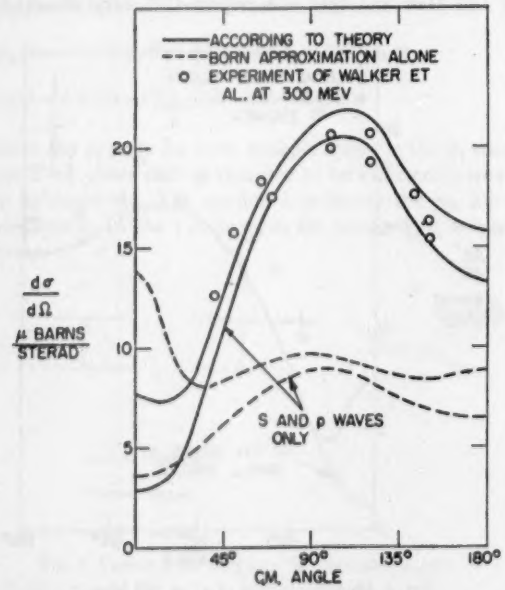

FIG. 6. Photoproduction of π^+ at 310 Mev.

[17] Fermi, Glicksman, Martin, and Nagle, Phys. Rev. **92**, 161 (1953).

Consider only s and p waves; then the scattering cross section is

$$d\sigma/d\Omega = A + B\cos\theta + C\cos^2\theta,$$

where

$$A/\lambda^2 = |a_{p3} - a_{p1}|^2 + |a_s|^2,$$
$$B/\lambda^2 = 2\,\mathrm{Re}\, a_s^*(2a_{p3} + a_{p1}),$$
$$C/3\lambda^2 = |a_{p3}|^2 + 2\,\mathrm{Re}\, a_{p3}^* a_{p1}.$$

The scattering amplitudes, $a_{l,2j}$, for the each of the three processes,

(a) $\pi^+ + p \to \pi^+ + p$,

(b) $\pi^- + p \to \pi^- + p$,

(c) $\pi^- + p \to \pi^0 + n$,

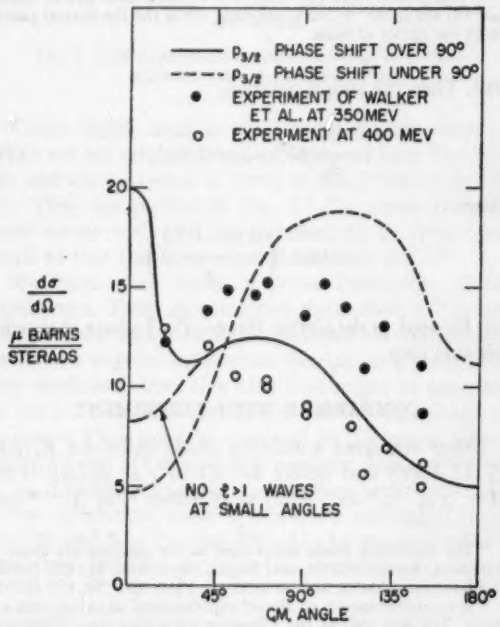
P3/2 PHASE SHIFT OVER 90°
P3/2 PHASE SHIFT UNDER 90°
EXPERIMENT OF WALKER ET AL. AT 350MEV
EXPERIMENT AT 400 MEV
NO ℓ>1 WAVES AT SMALL ANGLES
CM. ANGLE

FIG. 7. Photoproduction of π^+ at 375 Mev.

are defined in terms of the scattering amplitudes to the isotopic spin states ($a_{l,j,T} = e^{i\delta}\sin\delta$). These relations are

(a) $a_{lj} = a_{lj,T=\frac{1}{2}}$,

(b) $a_{lj} = \frac{1}{3}(a_{lj,T=\frac{1}{2}} + 2a_{lj,T=\frac{3}{2}})$,

(c) $a_{lj} = \frac{1}{3}\sqrt{2}(a_{lj,T=\frac{1}{2}} - a_{lj,T=\frac{3}{2}})$.

For case I it is found that the s-wave phase shifts can remain small. We assume then, to simplify the problem further, that the s wave $T=\frac{3}{2}$ phase shift is given by extrapolating from lower energy scattering work:

$$\delta_3 = 0.40 - 0.52(k/\mu). \qquad (12)$$

Table I gives the phase shifts chosen at 210 Mev to fit the π^- scattering data under these conditions. The

comparison with the data is given in Table II. Also shown in Table I are the phase shifts used to calculate the photoproduction cross sections at E_γ(lab) = 310 Mev and 375 Mev. These curves are shown in Figs. 6, 7. The s-wave phase shifts associated with case II are taken, essentially, from Fermi and Metropolis.[18,19]

The test between case I and case II is made in Fig. 7. The latter is quite unambiguously in disagreement with experiment, while the resonance case is in good agreement. It would be interesting to investigate the π^+ photoproduction experimentally at very small angles to further test this theory. The extension of $d\sigma^+/d\Omega(90°)$ in accordance with this result is given in Fig. 5. The π^0 photoproduction cross section at 375 Mev is given in Table III. The experimental neutral meson data[18] is rough, but is in agreement with a very large $\sin^2\theta$ term. There is also an underermined amount of forward asymmetry.

The π^+ total scattering cross section implied by the theory is shown in Fig. 8. Only the s and $p_{\frac{1}{2}}$ phase shifts are considered here. The experimental cross sections

TABLE I. Phase shifts used for calculations with $E_\pi > 135$ Mev.‖

	Angular momentum isotopic spin state					
	s		$p_{\frac{1}{2}}$		$p_{\frac{3}{2}}$	
	$T=\frac{1}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{3}{2}$
$E_\gamma=310$ Mev[a]	10	−20				60
$E_\pi=210$ Mev[b] Case I	20	−26	10	−10	−10	122
$E_\gamma=375$ Mev[c] Case I	10	−27				128
$E_\gamma=375$ Mev[c] Case II	0	−60				45

[a] Used in Fig. 6.
[b] Used in Table II.
[c] Used in Fig. 7 and Table III.

are, as yet, as rough as the theory!§ It will be recalled that the peak can be narrower or broader and can be shifted 10 Mev or so in either direction.

DISCUSSION

Qualitatively, why does the photoproduction behave as it does? The Born approximation (weak coupling limit) for photoproduction of π^+ is plotted at 310 Mev in Fig. 6. It is essentially isotropic. This is mainly due to the large s-state matrix element. Production to the high angular momentum states ($l>1$) consists of a low isotopic background and a large bump at extreme forward angles arising through interference with the s state matrix element. The asymmetry associated with the interference between the relatively small p-state matrix elements and the s state is over shadowed by the high angular momentum effects, but it is seen to be a small backward effect. The $p_{\frac{1}{2}}$ wave is too small to cancel the effect of the $p_{\frac{3}{2}}$ wave in this respect. In the Born approximation the neutral photoproduction is negligibly small.

When the meson is considered to interact with the nucleon in the final state there results a large increase in the $p_{\frac{3}{2}}$ matrix element. In the neighborhood of 300 Mev the shape for π^+ and π^0 photoproduction approaches that for a pure $M1$ transition to the $p_{\frac{3}{2}}$ state ($2+3\sin^2\theta$). For the π^0 the s state matrix element remains small. For the π^+ the s state contribution is reduced as a result of repulsion in the $T=\frac{3}{2}$ state, which also helps to wash out the forward bump associated with high angular momenta. The $p_{\frac{3}{2},\frac{3}{2}}$ matrix element essentially changes sign from the Born approximation as the $M1$ interaction becomes dominant,[10] so that both p states yield a backward asymmetry in interference with the s state. At higher energies the $s-p$ interference which is essentially proportional to the cosine of the difference between the phases of the s and p state *matrix elements*, vanishes and then becomes forward when the $p_{\frac{3}{2}}$ phase shift passes through 90°. The forward bump reappears as the $p_{\frac{3}{2}}$ state reintroduces a large forward component with which the high angular momenta interfere. In case II, as a result of the strong repulsion in the $T=\frac{3}{2}$, s state, the phase of the s-state matrix element is primarily determined by the contribution of the $T=\frac{1}{2}$ component which has a small phase. Thus the difference between the phases of $p_{\frac{3}{2}}$ and s matrix elements remains less than 90° and the asymmetry remains as it was at

TABLE II. π^- scattering from protons at 210 Mev, $A+B\cos\theta+C\cos^2\theta$ millibarns/sterad.

Process		A	B	C
$\pi^-\to\pi^-$	Theory	1.2	0.5	2.4
	Experiment[a]	1.56±0.34	0.50±0.47	2.14±1.09
$\pi^-\to\pi^0$	Theory	1.4	2.5	3.4
	Experiment[a]	0.84±0.70	1.94±0.73	5.56±2.31
$\pi^+\to\pi^+$	Theory	1.1	1.4	2.7

[a] See reference 17.

TABLE III. Photoproduction of π^0 from photons, $A+B\cos\theta+C\sin^2\theta$ microbarns/sterad.

E_γ(lab) Mev	A	B	C
265	3.7	1.1	3.9
375, Case I	4.1	1.2	11.3
375, Case II	6.9	6.0	8.5

[18] Unpublished report. In more detail, they find the $s_{\frac{1}{2}}$, $T=\frac{1}{2}$ phase shift to be small and negative and the $p_{\frac{1}{2}}$, $T=\frac{1}{2}$ phase shift to be about +30°. The latter would, it happens, have little effect if it were considered in the photoproduction curve Fig. 7. Other phase shift solutions have been found and studied by R. L. Martin, Bull. Am. Phys. Soc. **29**, No. 1, 28 (1954), and also by Glicksman, de Hoffmann and Metropolis (private communication).

[19] Homa, Goldhaber, and Lederman, Phys. Rev. **93**, 554 (1954); L. C. L. Yuan and S. J. Lindenbaum, Phys. Rev. **93**, 917 (1954).

§ This situation has changed with experiments reported by Grandey and Clark, Bull. Am. Phys. Soc. **29**, No. 1, 29 (1954); and J. Ashkin *et al.* In addition, Yuan and Lindenbaum are repeating their experiment with improved accuracy (private communication).

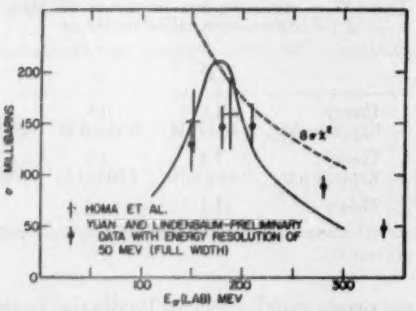

FIG. 8. Total cross section for scattering of positive mesons from protons. Only $p_{\frac{1}{2}}$ and $s_{\frac{1}{2}}$ phase shifts, taken from upper curve Fig. 3 and from Eq. (12), are considered. For experimental work see reference 19.

lower energies, in disagreement with experiment for positive mesons.

There is a rough consistancy between this phenomenological theory and experiment. The limitations of the theory are also quite evident. In the region near 310 Mev, where the assumption of energy independence should introduce only small errors, the agreement between theory and experiment is only within 10 or 20 percent. Of course absolute errors of this order may be present in the data, but this seems an unlikely explanation of some of the present difficulties. Improvement in the theory might be obtained by adjusting the coupling constant (i.e., see Fig. 5). Also the less important angular momentum isotopic spin states could be considered more fully. It is further seen, for example in $d\sigma^+/d\Omega(90°)$ (Fig. 5), how the theory breaks down completely at the high and low ends of the energy region. That the matrix elements should decrease in this region is indicated by examination of the Born approximation term;[8,30] and there is, perhaps, room for extension of the theory by making a detailed examination of the energy dependence of the various terms. Rapid changes are not indicated, however, and it seems clear that unless fairly rapid change of parameters with energy should be predicted, very good agreement with experiment would not be obtained. As our Born approximation term is only calculated in the weak coupling approximation, such difficulty is not surprising.

The author would like to thank Professor H. A. Bethe for his interest in this work.

[30] In the spirit of the present theory, the photoproduction reduces to the weak coupling limit as threshold is approached (see Fig. 4 with $G^2/4\pi=16$). This is in accord with the idea of N. M. Kroll and M. A. Ruderman, Phys. Rev. 93, 233 (1954). The cross section $d\sigma^+/d\Omega(90°)=6$ at about E_γ(lab)=175 Mev, recently reported by Bernardini (reference 16), would be fitted in this theory by a coupling constant of about 14.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

A Covariant Treatment of Meson-Nucleon Scattering

MAURICE M. LÉVY
École Normale Supérieure, Paris, France
(Received January 4, 1954)

A covariant equation for the meson-nucleon system is presented, in which the renormalization of divergent processes is carried out to all orders. A closed expression is given for their contribution to the wave function after renormalization, while the contribution coming from finite processes still involves a series expansion. Exact formulas are derived for the scattering phase shifts.

I. INTRODUCTION

RECENT experiments on pion-nucleon scattering[1] have made apparent the inadequacy of the Born approximation for the calculation of this process and the necessity of a theoretical analysis based on more elaborate methods.

Several attempts have been made[2] to analyze the data by means of the Tamm-Dancoff[3] nonadiabatic method, or an improved form of it.[4] Although this method seems to yield results which are in qualitative agreement with experiment, at least for the p wave, its defects are even more apparent here than in the treatment of nuclear forces.[5] A rapid calculation shows indeed that, even for low-energy scattering, high momenta play a decisive role in intermediate states, and that, consequently, the convergence of the interaction expansion can be expected to be very poor. Moreover, the main contribution to the scattering cross sections

[1] Barnes, Angell, Perry, Miller, Ring, and Nelson, Phys. Rev. 92, 1327 (1953); Bodansky, Sachs, and Steinberger, Phys. Rev. 93, 918 (1954); Anderson, Fermi, Martin, and Nagle, Phys. Rev. 91, 155 (1953).

[2] G. F. Chew, Phys. Rev. 89, 591 (1953); J. S. Blair and G. F. Chew, Phys. Rev. 90, 1065 (1953); S. Fubini, Nuovo cimento 10, 564 (1953); Dyson, Schweber, and Vissher, Phys. Rev. 90, 372 (1953); Sundaresan, Salpeter, and Ross, Phys. Rev. 90, 372 (1953); N. Fukuda, Proceedings of the International Conference of Kyoto, September, 1953 (unpublished).

[3] I. Tamm, J. Phys. U.S.S.R. 9, 449 (1945); S. M. Dancoff, Phys. Rev. 78, 382 (1950).

[4] F. J. Dyson, Phys. Rev. 91, 1543 (1953).

[5] M. M. Lévy, Phys. Rev. 88, 72, 725 (1952); A. Klein, Phys. Rev. 90, 1101 (1953).

comes from radiative corrections which cannot be handled correctly within the framework of the Tamm-Dancoff method.

These difficulties are partly removed if one uses, instead, a covariant nonadiabatic two-body integral equation, analogous to the one proposed by Salpeter and Bethe,[6] which can easily be extended to the present case. (As in Karplus *et al.*,[7] it will be called, in the following, the M.N. equation.) However, as in the case of quantum electrodynamics, renormalization—that is, the unambiguous elimination of unobservable divergent quantities—has still to be performed on the integral equation, using the methods of Feynman[8] and Dyson.[9]

In the present problem, however, special difficulties appear in the course of the renormalization process, which are essentially due to the fact that π mesons can play simultaneously the dual role of virtual field quanta and of real interacting particles. Mathematically, the difficulties arise through the fact that integral equations which possess perfectly finite kernels do not yield finite solutions, because they involve summation over a series of virtual processes, some of which include radiative effects. Partial solutions to these difficulties have been proposed by Karplus *et al.*[7] and Fubini.[10] In the treatment of these authors, however, renormalization is performed through a special device, the validity of which is limited to the lowest order, and it is not easily seen how the solution can be extended *to all orders*. Moreover, they work in terms of the Feynman two-body kernel, which is really convenient only when total cross sections need to be computed. When a phase-shift analysis of the experimental data is necessary, it is much easier to work in terms of the M.N. "wave function," as will be seen later.

The purpose of the present paper is to provide a framework within which scattering phase shifts and cross sections can be computed to all orders without renormalization difficulties. All the infinite diagrams are separated out, and a *closed* expression is given for their contribution to the wave function after renormalization. The calculation of the contribution coming from finite processes, however, still involves a series expansion, the convergence of which is not discussed here.[11]

The removal of divergences in the M.N. equation is most easily understood if, at first, only the two lowest-order diagrams (which are of the second order in the coupling constant) are included in the interaction kernel. This is done in Sec. II. The solution is presented, however, in such a form that the extension to all orders (Sec. III) is almost immediate. Exact expressions for the scattering phase-shifts are given in Sec. IV. In the handling of the M.N. "wave function," a good deal is implicitly used of a covariant theory of scattering, which starts from a relativistic four-dimensional wave equation, and is therefore the logical relativistic extension of the standard Faxén-Holtzmarck treatment[12] of the Schrödinger equation. This theory, which also yields a variational principle for the scattering phase-shifts, will be discussed in a subsequent paper.

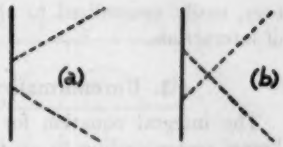

FIG. 1. The two basic diagrams of meson-nucleon scattering to the second order in the coupling constant.

The application of the formal results contained in the present paper to the analysis of low- and high-energy pion-nucleon scattering data will be presented later. The connection between meson-nucleon scattering and nuclear forces will also be discussed at that time.

II. COVARIANT EQUATION TO THE LOWEST ORDER IN THE COUPLING CONSTANT

In this section, only physical processes corresponding to an infinite number of iterations of the two basic diagrams of Fig. 1 will be considered. We shall call $K_M(x,y)$ and $K_N(x,y)$, respectively, the meson and nucleon propagation functions between two points x and y in space-time, their Fourier transforms being defined as follows:

$$K_M(x,y)=(2\pi)^{-4}\int K_M(p)\exp[ip(x-y)]d^4p,$$
$$K_N(x,y)=(2\pi)^{-4}\int K_N(p)\exp[ip(x-y)]d^4p. \tag{1}$$

These definitions imply therefore the following connections with the well known S_F and Δ_F functions, as defined by Dyson:[9]

$$K_M(p)=\tfrac{1}{2}\Delta_F(p)=-i(p^2+\mu^2)^{-1},$$
$$K_N(p)=-\tfrac{1}{2}S_F(p)=-i(i\gamma p+M)^{-1},$$

where μ and M are the meson and nucleon masses. The Feynman two-body kernel (see, for example, reference 6) will be written as $K(x,\xi;y,\eta)$.

For the sake of definiteness, we shall assume that we are dealing with a symmetrical mixture of pseudoscalar mesons with pseudoscalar coupling to the nucleons, writing therefore the interaction Hamiltonian as

$$H_{\text{int}}=iG\bar{\psi}\gamma_5\tau_k\psi\varphi_k, \tag{2}$$

where G is the coupling constant, ψ and φ_k the nucleon and meson fields, respectively. The theory can be, how-

[6] E. Salpeter and H. A. Bethe, Phys. Rev. **84**, 1232 (1951); M. Gell-Mann and F. Low, Phys. Rev. **84**, 350 (1951).
[7] Karplus, Kivelson, and Martin, Phys. Rev. **90**, 1072 (1953).
[8] R. P. Feynman, Phys. Rev. **76**, 749, 769 (1949).
[9] F. J. Dyson, Phys. Rev. **75**, 486, 1736 (1949).
[10] S. Fubini, Nuovo cimento **10**, 851 (1953).
[11] See, however, the concluding remarks of Sec. V.
[12] See, for example, N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford University Press, London, 1949), second edition, Chap. II.

ever, easily generalized to all "renormalizable" types of interaction.

1. Unrenormalized Equations

The integral equation for the Feynman two-body kernel corresponding to an infinite number of repetitions of Figs. 1(a) and 1(b) can be written as

$$K(x,\xi_i;\,y,\eta_j)=K_N(x,y)K_M(\xi_i,\eta_j)$$
$$-iG^2\int K_N(x,\xi')K_M(\xi_i,\xi_k')\gamma_5\tau_k K_N(\xi',\eta')\times\gamma_5\tau_l K(\eta',\eta_l';\,y,\eta_j)d^4\xi'd^4\eta'$$
$$-iG^2\int K_N(x,\xi')K_M(\xi_i,\eta_k')\gamma_5\tau_l K_N(\xi',\eta')\times\gamma_5\tau_k K(\eta',\xi_l';\,y,\eta_j)d^4\xi'd^4\eta', \quad (3)$$

where the isotopic spin indices have been written as subscripts to the meson variables. Equation (3) contains implicitly divergent higher-order effects such as, for example, those described by the two *reducible*[12] diagrams of Fig. 2. It leads to an unrenormalized equation for the wave function $\psi(x,\xi_i)$ defined as follows:

$$\psi(x,\xi_i)=\lim_{t_y,t_\eta\to-\infty}\int K(x,\xi_i;\,y,\eta_j)\gamma_4\psi_0(y,\eta_j)dy d\mathbf{n}, \quad (4)$$

where ψ_0 is the free wave function of the system,

$$\psi_0(x,\xi_i)=\lim_{t_y,t_\eta\to-\infty}\int K_N(x,y)K_M(\xi_i,\eta_j)\times\gamma_4\psi_0(y,\eta_j)dy d\mathbf{n}, \quad (5)$$

and which can be written

$$\psi(x,\xi_i)=\psi_0(x,\xi_i)$$
$$-iG^2\int K_N(x,\xi')K_M(\xi_i,\xi_k')\gamma_5\tau_k K_N(\xi',\eta')\times\gamma_5\tau_l\psi(\eta',\eta_l')d^4\xi'd^4\eta'$$
$$-iG^2\int K_N(x,\xi')K_M(\xi_i,\eta_k')\gamma_5\tau_l K_N(\xi',\eta')\times\gamma_5\tau_k\psi(\eta',\xi_l')d^4\xi'd^4\eta'. \quad (6)$$

2. Separation of the Divergences

We split the two-body kernel into two parts:

$$K=K_a+K_b, \quad (7)$$

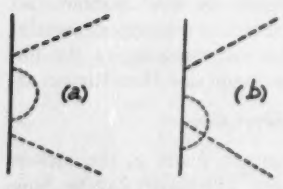

FIG. 2. Two reducible graphs of the meson-nucleon scattering matrix, which result from the combination, in high orders, of diagrams (a) and (b) of Fig. 1 and include divergent effects.

[12] A graph of the meson-nucleon scattering matrix is called *irreducible* if it is not possible to draw a line through it, cutting one nucleon line and one meson line only. Otherwise, it is called *reducible*.

where K_b contains only the *convergent* graphs *which result from the iteration of graph (1,b) alone* (examples are given in Fig. 3). The equation for K_b is simply obtained by suppressing the second term of the right-hand side of (3) and by writing K_b instead of K:

$$K_b(x,\xi_i;\,y,\eta_j)=K_N(x,y)K_M(\xi_i,\eta_j)$$
$$-iG^2\int K_N(x,\xi')K_M(\xi_i,\eta_k')\gamma_5\tau_l K_N(\xi',\eta')\times\gamma_5\tau_k K_b(\eta',\xi_l';\,y,\eta_j)d^4\xi'd^4\eta'. \quad (8)$$

The kernel function K_a contains all remaining divergent graphs, which essentially consist of three parts: a self-energy part which is an arbitrary combination of two types of diagrams, the general terms of which are represented in Figs. 4(a) and 4(b); and a vertex part on each side, belonging, respectively, to the general types described in Figs. 4(α) and 4(β). $K_a(x,\xi_i;\,y,\eta_j)$ can therefore be written formally as

$$K_a(x,\xi_i;\,y,\eta_j)=-iG^2\int K_N(x,x')K_M(\xi_i,\xi_k')\times\Gamma_b^{(\beta)}(\xi_k';\,x',y')\tau_k K_N'(y',x')\tau_l\Gamma_b^{(\alpha)}(\eta_l';\,x'',y'')\times K_N(y'',y)K_M(\eta_l',\eta_j)d^4x'd^4x''d^4y'd^4y''d^4\xi'd^4\eta'. \quad (9)$$

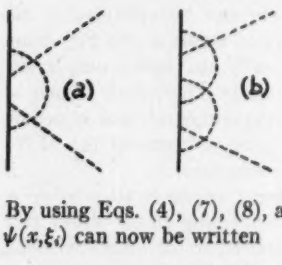

FIG. 3. Two reducible graphs corresponding to convergent processes and resulting from the iteration of diagram (b) of Fig. 1.

By using Eqs. (4), (7), (8), and (9), the wave function $\psi(x,\xi_i)$ can now be written

$$\psi(x,\xi_i)=\psi_a+\psi_b, \quad (10)$$

where ψ_b is the finite solution of the convergent integral equation

$$\psi_b(x,\xi_i)=\psi_0(x,\xi_i)-iG^2\int K_N(x,\xi')K_M(\xi_i,\eta_k')\times\gamma_5\tau_l K_N(\xi',\eta')\gamma_5\tau_k\psi_b(\eta',\xi_l')d^4\xi'd^4\eta', \quad (11)$$

and ψ_a is defined by the expression

$$\psi_a(x,\xi_i)=-iG^2\int K_N(x,x')K_M(\xi_i,\xi_k')\times\Gamma_b^{(\beta)}(\xi_k';\,x',y')\tau_k K_N'(y',x'')\tau_l\Gamma_b^{(\alpha)}(\eta_l';;\,x'',y'')\times\psi_0(y'',\eta_l')d^4x'd^4x''d^4y'd^4y''d^4\xi'd^4\eta'. \quad (12)$$

All the infinities are now concentrated in the vertex functions $\Gamma_b^{(\alpha)}$ and $\Gamma_b^{(\beta)}$, and in the modified nucleon propagation function K_N'. Our remaining task is to calculate these functions in closed form in terms of

K_b and ψ_b, and to renormalize them by means of the usual methods.

3. Calculation and Renormalization of $\Gamma_b^{(\alpha)}$ and $\Gamma_b^{(\beta)}$

Let us write for $\rho=\alpha,\beta$

$$\Gamma_b^{(\rho)}(\xi_i;x,y)=\gamma_b\delta(x-\xi)\delta(y-\xi)+\Lambda_b^{(\rho)}(\xi_i;x,y). \quad (13)$$

Calculating the successive contributions of the graphs contained in $\Lambda_b^{(\alpha)}$, one can write a power series expansion in G^2:

$$\Lambda_b^{(\alpha)}(\xi_i;x,y)=-iG^2\gamma_b\tau_k\Big\{K_N(x,\xi)K_M(x_k,y_k)$$
$$-iG^2\int K_N(x,\xi')K_M(x_k,\eta_k')\gamma_b\tau_l K_N(\xi',\eta')$$
$$\times\gamma_b\tau_k K_N(\eta',\xi)K_M(\xi_l',y_l)d^4\xi'd^4\eta'$$
$$+\cdots\Big\}\gamma_b\tau_i K_N(\xi,y)\gamma_b\tau_k, \quad (14)$$

and, by comparison with the corresponding expansion of $K_b(x,\xi_i;y,\eta_j)$, one obtains the following relation:

$$\Lambda_b^{(\alpha)}(\xi_i;x,y)=-iG^2\gamma_b\tau_k K_b(x,x_k;\xi,y_l)$$
$$\times\gamma_b\tau_i K_N(\xi,y)\gamma_b\tau_l. \quad (15)$$

Similarly, $\Lambda_b^{(\beta)}$ is expressed in terms of K_b by the equation

$$\Lambda_b^{(\beta)}(\xi_i;x,y)=-iG^2\gamma_b\tau_l K_N(x,\xi)$$
$$\times\gamma_b\tau_i K_b(\xi,x_i;y,y_k)\gamma_b\tau_k. \quad (16)$$

In order to perform the renormalization of these two vertex operators, we introduce first their Fourier transforms, which can be written, for $\rho=\alpha,\beta$, as

$$\Lambda_{k,i}^{(\rho)}(p,q)=\int\Lambda_b^{(\rho)}(\xi_i;x,y)$$
$$\times\exp[-ip(x-\xi)+iq(y-\xi)]d^4xd^4y. \quad (17)$$

The renormalized operators are then obtained by the usual method:

$$\Lambda_{k,i}^{*(\rho)}(p,q)=\Lambda_{k,i}^{(\rho)}(p,q)-\Lambda_{k,i}^{(\rho)}(p_0,p_0), \quad (18)$$

where p_0 is the energy momentum of a free nucleon satisfying the relation $(i\gamma p_0+M)=0$. In configuration space, Eq. (18) can be written as

$$\Lambda_b^{*(\rho)}(\xi_i;x,y)=\Lambda_b^{(\rho)}(\xi_i;x,y)$$
$$-\Lambda_{k,i}^{(\rho)}(p_0,p_0)\delta(x-\xi)\delta(y-\xi), \quad (19)$$

where $\Lambda_{k,i}^{(\rho)}(p_0,p_0)$, for $\rho=\alpha,\beta$, are two divergent constants which have to be calculated. Equation (17) gives

$$\Lambda_{k,i}^{(\rho)}(p_0,p_0)=\int\Lambda_b^{(\rho)}(\xi_i;x,y)$$
$$\times\exp[ip_0(y-x)]d^4xd^4y, \quad (20)$$

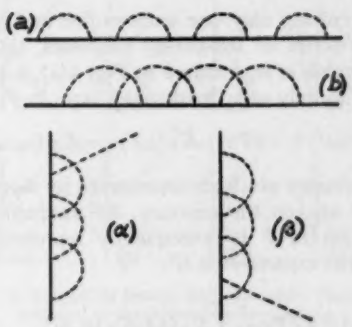

FIG. 4. General form of the divergent self-energy and vertex parts of diagrams corresponding to processes included in the kernel function K_a.

and consequently

$$\Lambda_{k,i}^{(\alpha)}(p_0,p_0)=-iG^2\gamma_b\tau_k\int K_b(x,x_k;\xi,y_l)\gamma_b\tau_i K_N(\xi,y)$$
$$\times\gamma_b\tau_l\exp[ip_0(y-x)]d^4yd^4x,$$
$$\Lambda_{k,i}^{(\beta)}(p_0,p_0)=-iG^2\gamma_b\tau_l\int K_N(x,\xi)\gamma_b\tau_i K_b(\xi,x_i;y,y_k) \quad (21)$$
$$\times\gamma_b\tau_k\exp[ip_0(y-x)]d^4yd^4x.$$

These relations can be simplified by expressing them in energy-momentum space, where K_b is defined by the equation

$$K_b(x,\xi_i;y,\eta_j)=(2\pi)^{-12}\int K_b^{(i,j)}(p,q;p'q')$$
$$\times\delta(p+q-p'-q')\exp[i(px+q\xi-p'y-q'\eta)]$$
$$\times d^4pd^4qd^4p'd^4q'. \quad (22)$$

The results are

$$\Lambda_{k,i}^{(\alpha)}(p_0,p_0)=-iG^2\gamma_k(2\pi)^{-8}\tau_k\int K_b^{(k,l)}$$
$$\times(p,p_0-p;q,p_0-q)\gamma_b\tau_i K_N(q)\gamma_b\tau_k d^4pd^4q,$$
$$\Lambda_{k,i}^{(\beta)}(p_0,p_0)=-iG^2\gamma_b\tau_k(2\pi)^{-8}\int K_N(p)\gamma_b\tau_i K_b^{(k,l)} \quad (23)$$
$$\times(p,p_0-p;q,p_0-q)\gamma_b\tau_l d^4pd^4q.$$

4. Calculation and Renormalization of $K_N'(x,y)$

The modified propagation function introduced in subsection (II,2) is, according to Dyson,[9] the solution of an integral equation of the form

$$K_N'(x,y)=K_N(x,y)+\int K_N(x,z)F(z,z')$$
$$\times K_N'(z',y)dzdz', \quad (24)$$

where $F(z,z')$ contains only the irreducible self-energy graphs, namely those which cannot be split into two

parts by cutting only one nucleon line and no meson line. The series of self-energy processes, the general term of which is represented in Fig. 4(a), corresponds therefore to only one (irreducible) term in F, namely,

$$F_1(z,z')=-iG^2\gamma_5\tau_k K_N(z,z')\gamma_5\tau_k K_M(z_k,z_k'). \quad (25)$$

The processes which are represented by diagrams like Fig. (4b) are, on the contrary, all irreducible. Their contribution to F is consequently represented by a power series expansion in G^2:

$$F_2(z,z')=(-iG^2)^2\gamma_5\tau_k\int K_N(z,\xi)K_M(z_k,\eta_k)$$
$$\times\gamma_5\tau_l K_N(\xi,\eta)\gamma_5\tau_k K_N(\eta,z')K_M(\xi_l,z_l')\gamma_5\tau_l d^4\xi d^4\eta$$
$$+(-iG^2)^3\gamma_5\tau_k\int K_N(z,\xi)K_M(z_k,\eta_k)\gamma_5\tau_m K_N(\xi,\eta)$$
$$\times\gamma_5\tau_k K_N(\eta,\xi')\gamma_5\tau_l K_M(\xi_m,\eta_m')\gamma_5\tau_m K_N(\eta',z')$$
$$\times K_M(\xi_l',z_l')\gamma_5\tau_l d^4\xi d^4\eta d^4\xi' d^4\eta'+\cdots. \quad (26)$$

A rapid inspection of this expansion shows that it is directly related to $K_b(z,z_k;\,z',z_l')$, through the equation

$$F_2(z,z')=-iG^2\gamma_5\tau_k[K_b(z,z_k;\,z',z_l')$$
$$-K_N(z,z')K_M(z_k,z_l')]\gamma_5\tau_l. \quad (27)$$

Combining (25) and (27) yields the relation

$$F(z,z')=F_1+F_2=-iG^2\gamma_5\tau_k K_b(z,z_k;\,z',z_l')\gamma_5\tau_l. \quad (28)$$

Expressing Eq. (24) in momentum space and using (22) and (28), we are led to a purely algebraic expression for $K_N'(p)$:

$$K_N'(p)=K_N(p)/[1-\Sigma(p)K_N(p)], \quad (29)$$

where we have introduced the notation

$$\Sigma(p)=-iG^2\gamma_5\tau_k\int K_b^{(k,l)}(q,p-q;\,q',p-q')$$
$$\times\gamma_5\tau_l d^4q d^4q'. \quad (30)$$

The only divergent quantity is evidently $\Sigma(p)$ which can be renormalized by the usual method (see reference 9):

$$\Sigma^*(p)=\Sigma(p)-\Sigma(p_0)$$
$$+\tfrac{1}{4}(i\gamma_\mu p_\mu+M)\left[i\gamma_\nu\frac{\partial\Sigma(p)}{\partial p_\nu}\right]_{p=p_0}, \quad (31)$$

where p_0 is again the energy-momentum of a free nucleon satisfying the relation: $i\gamma p_0+M=0$. The renormalized modified propagation function is then obtained by putting $\Sigma^*(p)$ instead of $\Sigma(p)$ in Eq. (29).

5. Final Expression of $\psi_a(x,\xi_i)$

Having obtained finite expressions for $\Gamma_{k,i}^{(\alpha)}$, $\Gamma_{k,i}^{(\beta)}$ and K_N', we are now in a position to calculate K_a and ψ_a in a closed finite form in terms of K_b and ψ_b:

$$K_a(x,\xi_i;\,y,\eta_j)=-iG^2\int K_N(x,\xi')K_M(\xi_i,\eta_k')$$
$$\times\{\gamma_5\tau_k[1-\Lambda_{0k}^{(\beta)}]\delta(\xi'-\eta')\delta(\xi''-\eta')$$
$$-iG^2\gamma_5\tau_m K_N(\xi',\eta')\gamma_5\tau_k K_b(\eta',\xi_m';\,\xi'',\xi''_k)\gamma_5\tau_k\}$$
$$\times K_N'(\xi'',\eta'')\{\gamma_5\tau_l[1-\Lambda_{0l}^{(\alpha)}]\delta(\eta''-\xi''')\delta(\xi'''-\eta''')$$
$$-iG^2\gamma_5\tau_r K_b(\eta'',\eta_r'';\,\xi''',\eta_s''')\gamma_5\tau_l$$
$$\times K_N(\xi''',\eta''')\gamma_5\tau_s\}K_N(\eta''',y)K_M(\xi''',\eta_j)$$
$$d^4\xi'\cdots d^4\xi'''d^4\eta'\cdots d^4\eta''', \quad (32)$$

where we have put $\Lambda_{0i}^{(\rho)}(p_0,p_0)=\gamma_5\Lambda_{0i}^{(\rho)}$, for $\rho=\alpha,\beta$. Equation (32) can be simplified by integrating over the δ functions and by making use of the integral equation (8) for $K_b(x,\xi_i;\,y,\eta_j)$, as well as another one which is completely equivalent to (8), namely,

$$K_b(x,\xi_i;\,y,\eta_j)=K_N(x,y)K_M(\xi_i,\eta_j)$$
$$-iG^2\int K_b(x,\xi_i;\,\xi',\eta_k')\gamma_5\tau_l K_N(\xi',\eta')$$
$$\times\gamma_5\tau_k K_N(\xi',y)K_M(\eta_l',\eta_j)d\xi'd\eta'. \quad (33)$$

The resulting expression for K_a is given by

$$K_a(x,\xi_i;\,y,\eta_j)=-iG^2\int[K_b(x,\xi_i;\,\xi',\xi_k')$$
$$-\Lambda_{0k}^{(\beta)}K_N(x,\xi')K_M(\xi_i,\xi_k')]\gamma_5\tau_k$$
$$\times K_N'(\xi',\eta')\gamma_5\tau_l[K_b(\eta',\eta_l';\,y,\eta_j)$$
$$-\Lambda_{0l}^{(\alpha)}K_N(\eta',y)K_M(\eta_l',\eta_j)]d\xi'd\eta'. \quad (34)$$

Making use of Eqs. (4) and (5), one obtains next an equation for $\psi_a(x,\xi_i)$:

$$\psi_a(x,\xi_i)=-iG^2\int[K_b(x,\xi_i;\,\xi',\xi_k')$$
$$-\Lambda_{0k}^{(\beta)}K_N(x,\xi')K_M(\xi_i,\xi_k')]\gamma_5\tau_k K_N'(\xi',\eta')$$
$$\times\gamma_5\tau_l[\psi_b(\eta',\eta_l')-\Lambda_{0l}^{(\alpha)}\psi_0(\eta',\eta_l')]d\xi'd\eta'. \quad (35)$$

This expression, which is a kind of inhomogeneous term to be added to ψ_b in order to get the exact wave function ψ, is actually a "contact" term since, in the center-of-mass system, it only involves the value of ψ_b at the origin. This leads to a particularly simple expression for the contribution of ψ_a to the S-matrix elements, as will be seen in Sec. IV.

III. RENORMALIZATION TO ALL ORDERS

The treatment given in Sec. II can now be extended to the irreducible processes corresponding to all orders in G^2 by very small modifications of the formalism.

Equations (7) and (10), through which the separation of divergences is performed, are still valid, and so are Eqs. (34) and (35) which express the renormalized functions K_a and ψ_a in terms of K_b and ψ_b. The latter quantities, however, obey now more general integral equations of the form

$$K_b(x,\xi_i;y,\eta_j)=K_N(x,y)K_M(\xi_i,\eta_j)+\int K_N(x,z)K_M(\xi_i,\xi_k)I_b(z,\zeta_k;z',\zeta_l')\times K_b(z',\zeta_l';y,\eta_j)dzdz'd\zeta d\zeta', \quad (36)$$

and

$$\psi_b(x,\xi_i)=\psi_0(x,\xi_i)+\int K_N(x,z)K_M(\xi_i,\zeta_k)\times I_b(z,\zeta_k;z',\zeta_l')\psi_b(z',\zeta_l')dzdz'd\zeta d\zeta', \quad (37)$$

where $I_b(z,\zeta_k;z',\zeta_l')$ is an interaction function which can be described by means of a series of graphs having the following properties:

(a) they are all irreducible (in the sense defined in reference 13);

(b) they contain no vertex or self-energy parts;[14]

(c) they are all finite and lead individually to finite expressions for K_b and ψ_b (we do not discuss here the convergence of I_b).

I_b can be expanded in a power series in G^2:

$$I_b=I_b^{(1)}+I_b^{(2)}+\cdots+I_b^{(n)}+\cdots, \quad (38)$$

where $I_b^{(n)}$ is proportional to G^{2n}. $I_b^{(1)}$ corresponds to diagram 1(b) and has therefore the expression

$$I_b^{(1)}=-iG^2\gamma_b\tau_l K_N(z,\zeta)\gamma_b\tau_l\delta(z-z')\delta(\zeta-\zeta'). \quad (39)$$

$I_b^{(2)}$ corresponds to only one fourth-order diagram, represented in Fig. 5(a), and is given by the relation

$$I_b^{(2)}=(-iG^2)^2\gamma_b\tau_m K_N(z,\zeta')\gamma_b\tau_l K_N(\zeta',\zeta)\times\gamma_b\tau_k K_N(\zeta,z')\gamma_b\tau_m K_M(z'_m,z_m). \quad (40)$$

$I_b^{(3)}$ corresponds to seven sixth-order diagrams, which are drawn in Fig. (5*b*,*c*,*d*,α,β,γ,δ), etc.

The above description makes it clear how the interaction function I_b has to be computed. The main point is that going to higher orders does not introduce any new divergence. All divergent processes have been, once and for all, separated out in K_a and ψ_a, and their renormalized contributions expressed in terms of the finite functions K_b and ψ_b.

It is perhaps worth mentioning that we have been concerned, so far, only in the removal of the special types of divergences introduced by the combination, in higher orders, of graph 1(a) with the other graphs of the meson-nucleon scattering matrix. There remain, of course, all the "normal" types of divergences which

[14] This means that they are also *irreducible* in the sense defined by Dyson (reference 9), whose definition is different from the one given in reference 13.

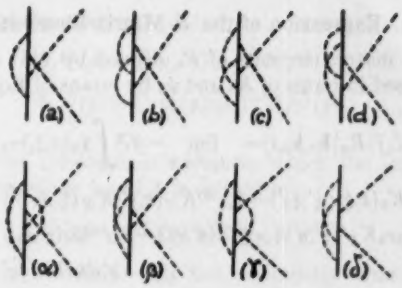

FIG. 5. Irreducible fourth- and sixth-order diagrams included in the interaction function I_b.

can be handled by standard methods: namely, in the final expressions for K_a, K_b, ψ_a, and ψ_b, all the K_N and K_M functions should be replaced by the exact modified propagation functions[15] K_N' and K_M', and γ_b by the vertex operator Γ_b. Finally, the contributions of processes involving meson-meson scattering should be renormalized by means of an additional nonlinear $\lambda\varphi^{*3}\varphi^3$ term in the interaction Hamiltonian[16] and added to the function I_b introduced in this section.

IV. CALCULATION OF SCATTERING PHASE-SHIFTS

The elements of the S matrix between two free states $\psi_0(x,\xi_i)$ and $\chi_0(x,\xi_j)$ are related to the Feynman two-body kernel by the relation

$$(\mathbf{k}_1',\mathbf{k}_2',j|S-1|\mathbf{k}_1,\mathbf{k}_2,i)=\lim_{\substack{t_x,t_\xi\to+\infty\\ t_y,t_\eta\to-\infty}}\int\bar{\chi}_0(x,\xi_j)\gamma_4\times\{K(x,\xi_j;y,\eta_i)-K_N(x,y)K_M(\xi_j,\eta_i)\}\times\gamma_4\psi_0(y,\eta_i)dxdyd\xi d\eta, \quad (41)$$

where the initial and final states are defined by the equations

$$\psi_0(x,\xi_i)=\psi_0^{(i)}(\mathbf{k}_1,\mathbf{k}_2)\exp[ik_1x+ik_2\xi],\qquad \chi_0(x,\xi_j)=\chi_0^{(j)}(\mathbf{k}_1',\mathbf{k}_2')\exp[ik_1'x+ik_2'\xi]. \quad (42)$$

Using Eqs. (4) and (5), putting $S-1=R=R_a+R_b$, we can write (41) as follows:

$$(\mathbf{k}_1',\mathbf{k}_2',j|R_a|\mathbf{k}_1,\mathbf{k}_2,i)=\lim_{t_x,t_\xi\to+\infty}\int\bar{\chi}_0(x,\xi_j)\times\gamma_4\psi_a(x,\xi_j)dxd\xi, \quad (43)$$

$$(\mathbf{k}_1',\mathbf{k}_2',j|R_b|\mathbf{k}_1,\mathbf{k}_2,i)=\lim_{t_x,t_\xi\to+\infty}\int\bar{\chi}_0(x,\xi_j)\gamma_4\times\{\psi_b(x,\xi_j)-\psi_0(x,\xi_j)\}dxd\xi. \quad (44)$$

[15] This modified propagation function K_N' (as well as the vertex operator Γ_b) should not be confused with the functions K_N', $\Gamma_b^{(\alpha)}$ and $\Gamma_b^{(\beta)}$ which we have introduced in Sec. II. These functions have been expressed, once and for all, in terms of the unmodified functions K_N, K_M and γ_b through Eqs. (13), (15), (16), (29), (30), and (36).

[16] A. Salam, Phys. Rev. 86, 731 (1952).

1. Expression of the R-Matrix Elements

The matrix elements of R_a defined by (43) can be expressed in terms of K_b and ψ_b by means of Eq. (35):

$$(\mathbf{k}_1',\mathbf{k}_2',j|R_a|\mathbf{k}_1,\mathbf{k}_2,i)=\lim_{t_a,t_\xi\to+\infty}-iG^2\int\bar{\chi}_0(x,\xi_j)\gamma_4\times[K_b(x,\xi_j;\xi',\xi_k')-\Lambda_{0k}{}^{(\beta)}K_N(x,\xi')K_M(\xi_i,\xi_k')]\times\gamma_\delta\tau_k K_N'(\xi',\eta')\gamma_\delta\tau_l[\psi_b(\eta',\eta_l')-\Lambda_{0l}{}^{(\alpha)}\psi_0(\eta',\eta_l')]\times dx d\xi d\xi' d\eta'. \quad (45)$$

This expression can be simplified further if one introduces the function $\bar{\chi}_b(y,\eta_i)$ defined by

$$\bar{\chi}_b(y,\eta_i)=\lim_{t_a,t_\xi\to+\infty}\int\bar{\chi}_0(x,\xi_j)\gamma_4 K_b(x,\xi_j;y,\eta_i)dx d\xi, \quad (46)$$

which is adjoint to that particular "wave function" of the system which, after an infinite time, transforms into the final free state $\chi_0(y,\eta_i)$. There exists obviously the relation

$$\bar{\chi}_0(y,\eta_i)=\lim_{t_a,t_\xi\to+\infty}\int\bar{\chi}_0(x,\xi_j)\gamma_4 K_N(x,y)\times K_M(\xi_j,\eta_i)dx d\xi, \quad (47)$$

and $\bar{\chi}_b(y,\eta_i)$ satisfies the integral equation

$$\bar{\chi}_b(y,\eta_i)=\bar{\chi}_0(y,\eta_i)+\int\bar{\chi}_b(z,\zeta_k)I_b(z,\zeta_k;z',\zeta_l')\times K_N(z',y)K_M(\zeta_l',\eta_i)dz dz' d\zeta d\zeta', \quad (48)$$

which can be obtained from Eq. (36) by multiplying on the right by $\bar{\chi}_0(x,\xi_j)$, integrating over the space components of x and ξ and letting the corresponding times go to infinity. Equation (45) becomes, then,

$$(\mathbf{k}_1',\mathbf{k}_2',j|R_a|\mathbf{k}_1,\mathbf{k}_2,i)=-iG^2\int[\bar{\chi}_b(\xi',\xi_k')-\Lambda_{0k}{}^{(\beta)}\bar{\chi}_0(\xi',\xi_k')]\gamma_\delta\tau_k K_N'(\xi',\eta')\gamma_\delta\tau_l\times[\psi_b(\eta',\eta_l')-\Lambda_{0l}{}^{(\alpha)}\psi_0(\eta',\eta_l')]d\xi' d\eta'. \quad (49)$$

It is now convenient to separate out the motion of the center of mass, by means of the transformation:

$$X=\alpha x+(1-\alpha)\xi,\qquad z=x-\xi, \quad (50)$$

where the parameter α expresses the well-known ambiguity in defining relativistically the position of the center of mass. We can write[17]

$$\psi_b(x,\xi_i)=\varphi_b{}^{(i)}(z)\exp(iP_\mu X_\mu), \quad (51)$$

where

$P_\mu=(P_0,\mathbf{P})$, $\mathbf{P}=\mathbf{k}_1+\mathbf{k}_2$, $P_0=E_{k_1}+\omega_{k_2}$, $E_{k_1}=(k_1^2+M^2)^{\frac{1}{2}}$, and

$$\omega_{k_2}=(k_2^2+\mu^2)^{\frac{1}{2}}.$$

Similarly, we have

$$\bar{\chi}_b(x,\xi_i)=\bar{\varphi}_b{}^{(i)}(z)\exp(-iP_\mu' X_\mu), \quad (52)$$

with $P_\mu'=(P_0',\mathbf{P}')$, $\mathbf{P}'=\mathbf{k}_1'+\mathbf{k}_2'$ and $P_0'=E_{k_1'}+\omega_{k_2'}$. The resulting expression of the R_a matrix elements is therefore

$$(\mathbf{k}_1',\mathbf{k}_2',j|R_a|\mathbf{k}_1,\mathbf{k}_2,i)=-iG^2(2\pi)^4[\bar{\varphi}_b{}^{(k)}(0)-\Lambda_{0k}{}^{(\beta)}\bar{\varphi}_0{}^{(k)}(0)]\gamma_\delta\tau_k K_N'(P)\gamma_\delta\tau_l\times[\varphi_b{}^{(l)}(0)-\Lambda_{0l}{}^{(\alpha)}\varphi_0{}^{(l)}(0)]\delta(\mathbf{P}-\mathbf{P}')\delta(P_0-P_0'). \quad (53)$$

In the same way, the matrix elements of R_b can be simplified to some extent by transforming them to the coordinate system (50). This gives:

$$(\mathbf{k}_1',\mathbf{k}_2',j|R_b|\mathbf{k}_1,\mathbf{k}_2,i)=(2\pi)^3\delta(\mathbf{P}-\mathbf{P}')\lim_{T\to+\infty}\{\exp[i(P_0-P_0')T]\}\times\int\bar{\rho}_0{}^{(j)}(\mathbf{z},0)[\varphi_b{}^{(i)}(\mathbf{z},0)-\varphi_0{}^{(i)}(\mathbf{z},0)]d\mathbf{z}, \quad (54)$$

where $\rho_0{}^{(j)}(\mathbf{z},0)\equiv\chi_0(\mathbf{k}_1',\mathbf{k}_2')\exp[i(\mathbf{k}_1'-\alpha\mathbf{P}')\cdot\mathbf{z}]$. In the derivation of Eq. (54), the adiabatic decoupling of the meson and nucleon fields when time goes to infinity has been implicitly assumed, since the limiting process t_x, $t_\xi\to+\infty$ has been performed on the assumption that $t_a-t_\xi\to 0$. Using the same procedure as Kita,[17] we introduce the definition

$$(\mathbf{k}_1',\mathbf{k}_2',j|R_b|\mathbf{k}_1,\mathbf{k}_2,i)=[\mathbf{k}_1',\mathbf{k}_2',j|R_b|\mathbf{k}_1,\mathbf{k}_2,i]\delta(\mathbf{P}-\mathbf{P}')\delta(P_0-P_0'), \quad (55)$$

and obtain the relation

$$\delta_+(P_0-P_0')[\mathbf{k}_1',\mathbf{k}_2',j|R_b|\mathbf{k}_1,\mathbf{k}_2,i]=(2\pi)^3\chi_0{}^{*(j)}(\mathbf{k}_1',\mathbf{k}_2')\times\int[\varphi_b{}^{(j)}(\mathbf{z},0)-\varphi_0{}^{(j)}(\mathbf{z},0)]\times\exp[-i(\mathbf{k}_1'-\alpha\mathbf{P}')\mathbf{z}]d\mathbf{z}, \quad (56)$$

where the identity

$$\lim_{T\to+\infty}\exp(iTx)\delta_+(x)=\delta(x)$$

has been used.

In the system where the center of mass is at rest ($\mathbf{k}_1=-\mathbf{k}_2=\mathbf{k}$, $\mathbf{k}_1'=-\mathbf{k}_2'=\mathbf{k}'$), this gives simply

$$\delta_+(P_0-P_0')[\mathbf{k}',-\mathbf{k}',j|R_b|\mathbf{k},-\mathbf{k},i]=(2\pi)^3\chi_0{}^{*(j)}(\mathbf{k}',-\mathbf{k}')\times\{\varphi_b{}^{(j)}(\mathbf{k}')-\varphi_0{}^{(j)}(\mathbf{k})(2\pi)^3\delta(\mathbf{k}-\mathbf{k}')\}, \quad (57)$$

where $\varphi_b{}^{(j)}(\mathbf{p})$ is the Fourier transform of $\varphi_b{}^{(j)}(\mathbf{z},0)$ and $\varphi_0{}^{(j)}(\mathbf{k})\equiv\psi_0{}^{(j)}(\mathbf{k},-\mathbf{k})$.

Scattering phase-shifts can then easily be obtained by means of a partial wave analysis of Eqs. (53) and (57).

2. Introduction of Three-Dimensional Wave Functions and Connection with the Tamm-Dancoff Method

The calculation of the R_a-matrix elements by means of Eq. (53) implies the knowledge of both functions $\bar{\chi}_b(y,\eta_i)$ and $\psi_b(x,\xi_i)$. It is perhaps more convenient to

[17] H. Kita, Progr. Theoret. Phys. Japan 7, 217 (1952).

obtain the corresponding phase-shifts through a direct analysis of the asymptotic behavior of the function $\psi_a(x,\xi_i)$ for equal times of the interacting particles. The connection with calculations based on the Tamm-Dancoff method will, at the same time, become more apparent.

Let us write, in the system of coordinates (50),

$$\psi_a(x,\xi_i)=\varphi_a^{(i)}(z)\exp(iP_\mu X_\mu), \tag{58}$$

and eliminate the X dependence by means of Eqs. (22) and (51). The resulting expression of $\varphi_a^{(i)}(z)$ is

$$\varphi_a^{(i)}(z)=-iG^2(2\pi)^{-4}\exp(-i\alpha Pz)\int K_N(p)K_M^{(i)}$$
$$\times(P-p)[R_b^{(i,k)}(p;P)-\Lambda_0^{(\beta)}(i,k)]\gamma_5\tau_k K_N'(P)\gamma_5\tau_l$$
$$\times[\varphi_b^{(l)}(0)-\Lambda_{0l}^{(\alpha)}\varphi_0^{(l)}(0)]\exp(ipz)d^4p, \tag{59}$$

where the following definition has been introduced:

$$(2\pi)^{-4}\int K_b^{(i,j)}(p,P-p;q,P-q)d^4q$$
$$=K_N(p)K_M^{(i)}(P-p)R_b^{(i,j)}(p;P), \tag{60}$$

$R_b^{(i,j)}(p;P)$ being a logarithmically divergent function which obeys the following integral equation:

$$R_b^{(i,j)}(p;P)=\delta_{ij}-iG^2(2\pi)^{-4}\gamma_5\tau_k\int K_N(u+p-P)$$
$$\times\gamma_5\tau_i K_N(u)K_M^{(k)}(P-u)R_b^{(k,j)}(u;P)du. \tag{61}$$

Since, according to Eqs. (23) and (60), $\Lambda_0^{(\beta)}(i,j)$ can be written in the form

$$\Lambda_0^{(\beta)}(i,j)=-iG^2(2\pi)^{-4}\gamma_5\tau_k\int K_N(u)\gamma_5\tau_i K_N(u)$$
$$\times K_M^{(k)}(P-u)R_b^{(k,j)}(u;p_0)du, \tag{62}$$

it is easily seen that

$$R_b^{*(i,j)}=R_b^{(i,j)}-\Lambda_0^{(\beta)}(i,j), \tag{63}$$

which appears on the right-hand side of (59), is a finite quantity. We now introduce the Fourier transform of $\varphi_a^{(i)}(z)$ through the equation

$$\varphi_a^{(i)}(z)=\exp[i(1-\alpha)Pz](2\pi)^{-4}\int\Phi_a^{(i)}(p)$$
$$\times\exp(ipz)dp, \tag{64}$$

and consequently,

$$\Phi_a^{(i)}(p)=-iG^2K_N(p+P)K_M^{(i)}(p)R_b^{*(i,k)}(p+P;P)$$
$$\times\gamma_5\tau_k K_N'(P)\gamma_5\tau_l[\varphi_b^{(l)}(0)-\Lambda_{0l}^{(\alpha)}\varphi_0^{(l)}(0)]. \tag{65}$$

Let us now split $\Phi_a^{(i)}(p)$ and $\Phi_b^{(i)}(p)$ (which is related to $\varphi_b^{(i)}(z)$ through an equation analogous to (64)), into their positive and negative energy components by means of the relations

$$\Phi_a^{(i)}(p)=\sum_\rho u_\rho(\mathbf{p})(\omega_p)^{-\frac{1}{2}}A_\rho^{(i)}(p),$$
$$\Phi_b^{(i)}(p)=\sum_\rho u_\rho(\mathbf{p})(\omega_p)^{-\frac{1}{2}}B_\rho^{(i)}(p), \tag{66}$$

where the Dirac-spinor $u_\rho(\mathbf{p})$, in which the index ρ takes the values 1 and 2, is defined by

$$[i\gamma\mathbf{p}-\gamma_4E_\rho(\mathbf{p})+M]u_\rho(\mathbf{p})=0, \tag{67}$$

with $E_1(\mathbf{p})=-E_2(\mathbf{p})=E_p$. Let us similarly write

$$\varphi_0^{(l)}(0)=\sum_\rho u_\rho(\mathbf{k})(\omega_k)^{-\frac{1}{2}}C_\rho^{(l)}. \tag{68}$$

Equation (65) can then be expressed in the form

$$A_\rho^{(i)}(p,p_0)=-\frac{G^2F_\rho^{(i)}(\mathbf{p},p_0;P_0)}{(\omega_p^2-p_0^2)[E_\rho(\mathbf{p})-p_0-P_0]}, \tag{69}$$

where the function $F_\rho^{(i)}$ is defined by

$$F_\rho^{(i)}(\mathbf{p},p_0;P_0)=\omega_p^{\frac{1}{2}}\omega_k^{-\frac{1}{2}}\sum_{\rho'}\bar{u}_\rho(\mathbf{p})R_b^{*(i,k)}$$
$$\times(\mathbf{p},p_0+P_0;0,P_0)\gamma_5\tau_k K_N'(0,P_0)$$
$$\times\gamma_5\tau_l u_{\rho'}(\mathbf{k})B_{\rho'}^{*(l)}, \tag{70}$$

with $p_0=-ip_4$, the finite constant $B_\rho^{*(i)}$ being expressed as

$$B_\rho^{*(i)}=\left\{i(2\pi)^{-4}\sum_{\rho'}\int u_\rho^*(\mathbf{k})u_{\rho'}(\mathbf{p})\right.$$
$$\left.\times\omega_k^{\frac{1}{2}}\omega_p^{-\frac{1}{2}}B_{\rho'}^{(i)}(\mathbf{p},p_0)d\mathbf{p}dp_0\right\}-\Lambda_{0l}^{(\alpha)}C_\rho^{(l)}. \tag{71}$$

In order to obtain the three-dimensional Fourier-component of the wave function corresponding to equal times of the particles, we must[5] calculate the quantity:

$$a_\rho^{(i)}(p)=-\frac{1}{2\pi i}\int_{-\infty}^{+\infty}A_\rho^{(i)}(\mathbf{p},p_0)dp_0.$$

While doing the integration over p_0, however, we can take advantage of the fact that we only need the *asymptotic* form of the Fourier-transform $f_\rho^{(i)}(\mathbf{r})$ of $a_\rho^{(i)}(\mathbf{p})$, so that only the poles exhibited on the right-hand side of (69)—and not those eventually existing in $F_\rho^{(i)}$—will contribute to the integral. The result is

$$a_\rho^{(i)}(\mathbf{p})\sim\frac{G^2\epsilon_\rho F_\rho^{(i)}(\mathbf{p},-\epsilon_\rho\omega_p;P_0)}{\omega_p+\epsilon_\rho[E_\rho(\mathbf{p})-P_0]}, \tag{73}$$

with $\epsilon_1=+1$, $\epsilon_2=-1$, $E_\rho(\mathbf{p})=\epsilon_\rho E_p$. Only the positive energy component ($\rho=1$) of $a_\rho^{(i)}(\mathbf{p})$ does not vanish at infinity. Introducing polar coordinates (p,ϑ,φ) and (r,Θ,Φ) in momentum and configuration spaces, respectively, we can analyze $a_1^{(i)}(\mathbf{p})$ and $f_1^{(i)}(\mathbf{r})$ into partial waves:

$$a_1^{(i)}(p)=\sum a_{1,lm}^{(i)}(p)Y_{l,m}(\vartheta,\varphi),$$
$$f_1^{(i)}(r)=\sum f_{1,lm}^{(i)}(r)Y_{l,m}(\Theta,\Phi). \tag{74}$$

where the sums on the right-hand sides involve also a summation over spin-eigenfunctions. We then get for the asymptotic form of $f_{l,lm}^{(i)}(r)$

$$f_{l,lm}^{(i)}(r)\sim\frac{i^l}{2\pi^2}G^2\int_0^\infty\frac{j_l(pr)F_{l,lm}^{(i)}(p,-\omega_p;P_0)}{\omega_p+E_p-P_0}\times p^2dp, \quad (75)$$

where $j_l(x)=(\pi/2x)^{-\frac{1}{2}}J_{l+\frac{1}{2}}(x)$. Since the main contribution to the asymptotic form comes from the pole $p=k$ (defined by $P_0=\omega_k+E_k$), we can write

$$f_{l,lm}^{(i)}(r)\sim\frac{i^l}{\pi^2}G^2\frac{E_k\omega_k}{P_0}F_{l,lm}^{(i)}(k,-\omega_k;P_0)\int_0^\infty\frac{j_p(pr)p^2dp}{p^2-k^2}$$

$$\sim\frac{i^l}{2\pi}G^2\frac{kE_k\omega_k}{P_0}F_{l,lm}^{(i)}(k,-\omega_k;P_0)g_l(kr), \quad (76)$$

with $g_l(x)=(\pi/2x)^{-\frac{1}{2}}J_{-(l+\frac{1}{2})}(x)$. Because of the definition of $\varphi_0^{(i)}(0)$ in Eq. (68), Eq. (76) leads to the following expression for the contribution of ψ_a to the phase shift:

$$\tan\delta_{lm}^{(i)}(a)=\frac{(-1)^{\frac{1}{2}}G^2k}{4\pi^{\frac{1}{2}}}\frac{E_k\omega_kF_{l,lm}^{(i)}(k,-\omega_k;P_0)}{P_0(2l+1)^{\frac{1}{2}}C_1^{(i)}}. \quad (77)$$

The same method can be used to calculate the contribution of ψ_b to the scattering phase shifts. In this case, however, the reduction of integral equation (37) to a three-dimensional form involves some manipulations, which can be based, for example, on the iteration method which has been worked out for the nuclear two-body problem[8] and extended to pion-nucleon scattering by Deser and Martin.[18]

A final remark should be made about the factor $\gamma_b\tau_bK_N'(0,P_0)\gamma_\beta\tau_\beta$ on the right-hand side of Eq. (70). If the $K_b^{(k,l)}$ function which appears on the right-hand side of Eq. (30) were replaced by its zero-order approximation $K_N(q)K_M(p-q)\delta_{kl}$, this factor would just be the damping coefficient calculated by Brueckner, Gell-Mann, and Goldberger.[19] However, since the exact kernel does probably not have the same dependence on momenta and isotopic spin, the resulting effect might well be entirely different. The results of the present section provide, in fact, a method to ascertain to what extent the damping effects predicted by Wentzel[20] are actually present in pion-nucleon scattering.

V. CONCLUDING REMARKS

It might appear, at first sight, rather surprising that meson-nucleon scattering can be formulated covariantly in a much more compact form than other two-body problems. The main reason is probably that the meson-nucleon system is not properly a two-body system, but should be more appropriately described as a "radiative one-body system." It is this reason which makes the writer feel that it might be possible to formulate the problem in an even more compact way than has been done in the present paper. There might exist, for example, an integral equation for the finite interaction function I_b introduced in Sec. III, which would enable one to solve the problem in a completely closed form. So far, however, only integral equations which take into account parts of the diagrams contained in I_b have been obtained.

Another problem is, of course, to determine to what extent the predictions of the theory agree with experiment. This problem is partly related to the preceding one, since the convergence of the series of diagrams described by I_b for any physically acceptable value of G^2 might well be doubtful. The agreement with experiment also depends on the compared magnitude of the "damping" or "resonance" effects (depending on their sign) in the contributions of ψ_a and ψ_b to the scattering phase shifts. It is likely that the influence of renormalization, which leads to a completely different treatment of these two parts of the wave function, will be felt there very strongly.

The author is indebted to Professor R. E. Marshak for many stimulating discussions.

Additional note: After the completion of this work, we have received, through a private communication, a summary of the results obtained by D. Ito and H. Tanaka, which are similar to those of S. Fubini (reference 10).

Note added in proof.—In order to take properly into account the difficulties of overlapping divergences, the renormalization prescriptions (18) and (31) have to be replaced, respectively, by the following:

$$\Gamma_{b,i}^{*(\rho)}(p,q)=\frac{\Gamma_{b,i}^{(\rho)}(p,q)}{1+\Lambda_{0i}^{(\rho)}}, \quad \text{(a)}$$

$$\Sigma^*(p)=\Sigma_1(p)-\Sigma_1(p_0)-(p-p_0)\left[\frac{\partial}{\partial p}\Sigma_1(p)\right]_{p=p_0}, \quad \text{(b)}$$

where $\sum_1(p)$ is defined by the equation

$$\Sigma_1(p)=\frac{\Sigma(p)}{1+\Lambda_{0k}^{(\alpha)}+\Lambda_{0l}^{(\beta)}}. \quad \text{(c)}$$

These prescriptions preserve the main conclusion of the present paper, namely that the vertex operators $\Gamma_b(\rho)$ and the modified propagation function $K_N'(p)$ can be renormalized in closed form and expressed in terms of the finite functions ψ_b and K_b. Some of the equations of Sec. IV have to be modified accordingly. For example, Eq. (53) becomes

$$(\mathbf{k}_1',\mathbf{k}_2',j\,|\,R_a\,|\,\mathbf{k}_1,\mathbf{k}_2,j)=-iG^2(2\pi)^4\frac{\delta_b^{(k)}(0)}{1+\Lambda_{0k}^{(\beta)}}$$

$$\times\gamma_b\tau_bK_N'(P)\gamma_\beta\tau_\beta\frac{\phi_b^{(l)}(0)}{1+\Lambda_{0l}^{(\alpha)}}\delta(\mathbf{P}-\mathbf{P}')\delta(P_0-P_0').$$

Proofs of prescriptions (a), (b), and (c) will be given in a subsequent note.

[18] S. Deser and P. C. Martin, Phys. Rev. **90**, 1075 (1953).
[19] Brueckner, Gell-Mann, and Goldberger, Phys. Rev. **90**, 476 (1953).
[20] G. Wentzel, Phys. Rev. **86**, 802 (1952).

The Continuous Spectrum Approximation in Quantum Statistics

PETER T. LANDSBERG*
Randal Morgan Laboratory of Physics, University of Pennsylvania, Philadelphia, Pennsylvania
(Received December 14, 1953)

A class O of systems of noninteracting quantum-mechanical particles is defined, so that it is often justifiable to approximate a given system of particles as a system of class O. A "limit L" is defined as the limiting case which arises of the number of particles N and the volume V of the system are both allowed to tend to infinity so as to keep the ratio N/V a finite and nonzero constant. It is shown that if one wishes to calculate the mean value of an extensive variable, as averaged over a canonical or over a grand canonical ensemble, for a system of class O in the limit L, one can apply the continuous spectrum approximation directly to the particle quantum states (excepting, possibly, the states which belong to the lowest energy level) without using a limiting process.

BY the continuous spectrum approximation in quantum statistics we usually mean that a summation over quantum states is replaced by a corresponding integration, using a suitable continuous density-of-states function, which gives the number of quantum states per unit energy interval as a function of energy. If such a density-of-states function is known, this procedure is very easily handled, and it therefore often represents the most direct attack on a given problem. But the results obtained in this way are not exact if a part of the spectrum, or the whole spectrum, is discrete. This detracts from the merit of the procedure. We shall show, however, that

the results obtained by the continuous spectrum approximation often apply with sufficient accuracy to a certain limiting case. (1)

We first elucidate what limiting case is referred to in (1). This denotes the "limit L" in which the total number of particles $\bar{N}$, as averaged over a grand canonical ensemble, and the volume V of each system of the ensemble are each allowed to tend to infinity so as to keep the ratio $\bar{N}/V$ a finite and nonzero constant.

We elucidate next what is meant by the word "often" in (1). It indicates that attention is confined here to a certain "class O" of systems. This is the class of all systems of noninteracting quantum-mechanical particles, such that each satisfies the following requirements: (i) There is an infinity of particle quantum states whose particle energies are E_1, $E_2, \cdots$; the labeling of the states is arranged so that $E_j \leq E_{j+1}$ for all j; (ii) E_j is finite if j is finite; (iii) all energy levels with finite j behave like $E_j = E_0 + b_j V^{-\theta}$ for large volumes V, where b_j is independent of the volume, and E_0 is the value to which all those E_j's tend in the limit L for which j is finite; (iv) for large volumes there exists a finite positive integer M, such that a continuous spectrum approximation is valid for $j > M$, and the jth quantum state is give by $E_j = E_0 + cV^{-\theta} j^{\tau}$ for $j > M$. The quantities c, θ, τ are positive constants, and j is a continuous index for $j > M$. M must exceed the degeneracy ω_1, of the lowest particle energy level E_1 for the system of interest when its volume is finite. Under (iv) one might equivalently specify the density-of-states function:

$$N(\epsilon) = \frac{1}{\tau}\left(\frac{V^{\theta}}{c}\right)^{1/\tau} (\epsilon - E_0)^{(1/\tau)-1}. \tag{2}$$

It is known[1] that for particles in a w-dimensional rectangular box, these assumptions are a good approximation, provided one chooses

$$\theta = \tau = \frac{2}{w}, \quad c = \left[\frac{\Gamma(1+\frac{1}{2}w)}{2^{w-2}}\right]^{2/w} \frac{h^2}{2\pi m}, \quad (w = 1, 2, \cdots). \tag{3}$$

We shall find it convenient to divide the energy spectrum of a particle in a system of class O into three groups of quantum states: group (A) contains the states belonging to the lowest energy level E_1; group (B) contains the other quantum states up to, and including, the Mth quantum state; group (C) contains the remaining quantum states (for which $j > M$).

In order to prove our assertion (1), we now consider a system of class O in the limit L. Let F be an extensive property of the system, $F(E_j)$ the contribution to F which is made by a particle in the jth quantum state. We shall assume that $F(E_j)$ is nonnegative. It will also be assumed to be finite, provided the system is of finite size, and provided also j is finite. The mean value of F, as averaged over a grand canonical ensemble and expressed per unit volume of the system of interest, is ($\alpha + E_0/kT$ is the usual chemical potential)

$$\frac{\bar{F}}{V} = \sum_j \frac{F(E_j)/V}{\exp[(E_j - E_0)/kT - \alpha] \pm 1}. \tag{4}$$

The summation is over all particle quantum states, and the top sign applies to fermions, the bottom sign to bosons. Equation (4) is exact;[2] the introduction of E_0 represents a slight generalization of the usual formulas.

We must now consider three cases. (a) In the case of fermions each term on the right-hand side of (4) is

* On leave of absence from the University of Aberdeen, Aberdeen, Scotland.

[1] This is a simple generalization of the usual results.

[2] R. C. Tolman, *Principles of Statistical Mechanics* (Oxford University Press, London, 1938), p. 512.

finite if j is finite, whatever the temperature. By assumption (iii) and the finiteness of M, the first M terms on the right-hand side of (4) can therefore be neglected in the limit L. (b) In the case of a Bose-Einstein gas above its condensation temperature the same considerations apply since $\alpha<(E_1-E_0)/kT$ in this case. The condensation temperature, if it exists, is here defined as that temperature T_c such that the fraction of the mean total number of particles which is, in the limit L, in the lowest energy level, is zero for $T>T_c$, and positive for $T<T_c$.[3] (c) In the case of a Bose-Einstein gas below its condensation temperature, it is known[3] that $0<\theta=\tau<1$, and that $\alpha\leq(E_1-E_0)/kT$. Hence,

$$\frac{\bar{F}}{V}-\frac{\omega_1}{V}\frac{F(E_1)}{\exp[(E_1-E_0)/kT-\alpha]-1}=\sum_{j=\omega_1+1}\frac{F(E_j)/V}{\exp[(E_j-E_0)/kT-\alpha]-1}. \quad (5)$$

For each term on the right-hand side of (5),

$$0\leq\frac{F(E_j)/V}{\exp[(E_j-E_0)/kT-\alpha]-1}\leq\frac{F(E_j)/V}{\exp[(E_j-E_1)/kT]-1}. \quad (6)$$

If j is finite, the right-hand side of (6) becomes, for large enough volumes, in virtue of assumptions (ii) and (iii),

$$\frac{F(E_j)}{V}\frac{kTV^\theta}{b_j-b_1},$$

and tends to zero in the limit L (since $\theta<1$). It follows by (6) that each term on the right-hand side of (5) tends to zero in the limit L, provided its value of j is finite. By assumption (i) this leaves an infinity of terms which have not been proved to vanish. Combining the results obtained, it has been shown that the states of the group (B) of quantum states can be neglected on the right-hand side of (4) in the limit L in all three cases, and, in addition, the states of group (A) can be neglected in cases (a) and (b). By assumptions (i) and (iv) this still leaves an infinity of quantum states in group (C), which contribute

$$\frac{1}{V}\int_{\varepsilon_{M+1}}^{\infty}\frac{F(E)N(E)dE}{\exp[(E-E_0)/kT-\alpha]\pm1} \quad (7)$$

to the right-hand side of (4). In all important cases $\theta=\tau$ and $N(E)$ is then by Eq. (2) proportional to the volume. (7) is therefore nonzero in the limit L, and (4) becomes

$$\lim\left(\frac{\bar{F}}{V}\right)=\lim\left\{\frac{\omega_1F(E_1)/V}{\exp[(E_1-E_0)/kT-\alpha]\pm1}\right\}+\int_{\varepsilon_0}^{\infty}\frac{F(E)N'(E)dE}{\exp[(E-E_0)/kT-\alpha]\pm1}. \quad (8)$$

We have here put $N'(E)=N(E)/V$. The first term need be retained only for a Bose-Einstein gas below its condensation temperature T_c, when its value can easily be shown to be[3] $[F(E_0)\bar{N}/V][1-(T/T_c)^{1/2}]$. The last term in (8) is just what one would have obtained if one had applied the continuous spectrum approximation to the whole of the particle energy spectrum in the first place.

Our result will now be stated. Let the value of an extensive variable, which is exhibited by a system of class O and which is positive and finite for finite systems, be averaged over a grand canonical ensemble, using two different methods. In method I finite systems are treated, using the continuous spectrum approximation for the whole spectrum of particle quantum states, and the density of states function appropriate for high quantum numbers. In method II the continuous spectrum approximation is used for the states of group (C), but otherwise the average is calculated exactly for the limit L. Our theorem then states that both methods yield the same result, provided only that in the case of a Bose-Einstein gas at, or below its condensation temperature, the lowest energy level is treated separately, and E_1 is replaced by E_0, in method I.

If the averaging procedure is based on a canonical ensemble, an identical theorem holds, except that an additional approximation must be stipulated for method I. It consists in approximating the ratio $\bar{n}(N+1, j)/\bar{n}(N,j)$ by unity. $\bar{n}(N,j)$ denotes here the mean occupation number of the jth particle quantum state as averaged over a canonical ensemble, when the total number of particles in the system is N. A proof of this second theorem follows directly from the first theorem and the relation between the grand canonical ensemble and the canonical ensemble, which has been discussed elsewhere.[4]

Lastly, as a special case of a theorem proved recently,[4] we have a third theorem which states that in method II the same result is obtained whether one averages over a canonical or a grand canonical ensemble. This is a consequence of the fact that the mean occupation numbers are the same in the limit L. It follows that the four methods of calculating the mean value of an extensive variable, which are the subject of the first two theorems, all lead to the same result.

We wish to call attention to three significant points. In the first place the assumptions of method I are just those which are often made. Method II is sufficiently exact for systems of class O, and our theorem therefore enables one to interpret the results of method I as sufficiently exact limiting values. The quantity of great mathematical importance is then seen to be the density of quantum states in group (C), i.e., for high quantum numbers. Secondly, while method II is sufficiently exact, an approximation occurs in setting up the problem, for the whole energy spectrum of the

[3] P. T. Landsberg, Proc. Cambridge Phil. Soc. **50**, 65 (1954).

[4] P. T. Landsberg, Phys. Rev. **93**, 1170 (1954).

(finite) system of interest is in fact discrete. One must therefore choose both V and M large enough to validate the treatment of the system of interest as a system of class O. Lastly, the theorems have the property of being independent of the precise value of M. This means that the group (B) of states, which has been shown to have no effect on the mean values of extensive variables, can be chosen to be as large as is compatible with the finiteness of M. The reason for this result can be explained as follows: as the volume and the total number of particles in the system are both increased, the number of quantum states which contribute materially to the average value $\bar{F}$ of an extensive variable increases up to and beyond any given finite integer. The neglect of a finite number of quantum states represents therefore a negligible error, provided only the occupation number of each state increases sufficiently slowly as the average total number of particles, $\bar{N}$, is increased. As has been seen, $\bar{n}_j/V$ does in fact tend to aero in the limit L for all states of group (B). For this reason, in the limit L, almost the whole contribution to $\bar{F}$ can be regarded as coming from those states for which the continuous spectrum approximation is satisfactory when the system is finite.

We have previously explained[3] why similar results are obtained in the limit L for the condensation temperature and the mean occupation number of a perfect Bose-Einstein gas above or below its condensation temperature in the following four cases: when the analysis is based on a grand canonical ensemble (for a system of class O in the limit L), when it is based on a canonical ensemble (again for a system of class O in the limit L), or when the continuous spectrum approximation is used in either of the two procedures which we have here called "Method I." These results follow from the general theorems we have proved above. They are therefore characteristic of the quantum statistics of perfect gases in general, and are not peculiar only to Bose-Einstein systems. It will be shown elsewhere that these theorems can be used to give a restricted quantum statistical proof of the third law of thermodynamics.

The four methods of calculation which we have discussed can be carried through independently of the method of steepest descents. This is desirable, since the validity of this method, *as usually formulated*, is doubtful for a Bose gas at low temperatures. We showed in a previous paper[3] that in this case (i) the first order correction term is to the mean occupation numbers do not diverge (contrary to what was previously believed), and (ii) *to the first order correction terms in the usual method of steepest descents* the partition function Z_N does not diverge.[5] However, both the main term and the first order correction term to $\log Z_N$ diverge, though in such a manner that their sum is again finite. We did not emphasize in our previous paper, but wish to do so here in order to remove all possibility of misunderstanding, that in order to assess the reliability of the usual way of applying the method of steepest descents to a Bose gas at low temperatures, the findings (i) and (ii), which have just been mentioned, force one to carry out a more general investigation of the higher order correction terms to $\log Z_N$ and to the mean occupation numbers. Only then can one be sure if and why the method fails in this case, and how it can be improved. The results which have been established in this paper, are independent of the outcome of such an investigation.

[5] Our statement on p. 74 of reference 3, "None of the first order correction terms diverges" should be amended in this sense. I am grateful to Dr. R. B. Dingle for drawing my attention to this point.

PHYSICAL REVIEW VOLUME 94, NUMBER 2 APRIL 15, 1954

The Classical Equations of Motion of Point Particles in Neutral Meson Fields

FREDERIC R. CROWNFIELD, JR.,* AND PETER HAVAS
Department of Physics, Lehigh University, Bethlehem, Pennsylvania

(Received August 10, 1953; revised manuscript received December 21, 1953)

The special relativistic equations of motion of point particles interacting with neutral scalar or vector meson fields are redetermined by applying a method originally developed by Infeld and Wallace for the electromagnetic case, which is based on the linearized equations of general relativity. This method, which is free of ambiguity, leads to the equations obtained previously by several authors by the method originated by Dirac in electrodynamics, but differs from some of the results of Majumdar and Apte, who used a simplified version of the Infeld-Wallace method, which did not have recourse to general relativity.

I. INTRODUCTION

FROM considerations of classical special relativity as well as of quantum theory it appears to be advantageous to treat elementary particles as mathematical points. Several methods have recently been suggested to avoid the infinities associated with point singularities in classical field theory, and to derive the equations of motion of such singularities from field theoretical considerations.

The method originally developed by Dirac[1] in

* Holder for part of this work of a George Gowen Hood Fellowship, which is gratefully acknowledged. Present address: Department of Physics, North Carolina State College, Raleigh, North Carolina.

[1] P. A. M. Dirac, Proc. Roy. Soc. (London) A167, 148 (1938).

electrodynamics requires that the conservation laws for energy and momentum valid for the field alone should persist in the presence of singularities. This method has been applied to the motion of simple poles of neutral vector and scalar meson fields by Bhabha[2] and Harish-Chandra[3] using retarded fields. The equations of motion for these cases using half-retarded, half-advanced fields were given in a recent paper,[4] which also contains references to further developments of this method and an action-at-a-distance formulation of the theory.

Dirac's method does not determine the equations of motion unambiguously. In the first place, the requirement of conservation laws does not determine the form of the energy-momentum vector of the particle uniquely; the equations given in the above references are only the simplest of several possible ones. Second, as exemplified by the case of fields of zero rest mass,[5] the field alone may obey more conservation laws than we can expect for the system of particle and field; there is, however, no mathematical reason to require one rather than another such conservation law to persist for the system.

A possible way to remove these ambiguities is to require the equations of motion of the special theory of relativity to fit within the framework of general relativity. One method of obtaining such equations, based on the work of Mathisson,[6] has been indicated by Lubanski.[7] Another method, suggested by the work of Einstein, Infeld, and Hoffmann,[8] has been developed by Infeld and Wallace.[9]

This method (hereafter called the complete IW method) was applied by these authors to the electromagnetic case. They also described a related method (hereafter called the simplified IW method), which is based on the conservation laws rather than on general relativity, and which thus again does not determine the equations of motion uniquely. This simplified method was used by Majumdar and Apte[10] to calculate the equations of motion of simple poles of neutral meson fields. They obtained results which in the retarded vector case are in agreement with those of Bhabha;[2] in the retarded scalar case they do not agree with those of Harish-Chandra,[3] however.

The present study was undertaken to obtain an unambiguous answer to the problem of the correct form of the field theoretical equations of motion of point particles in neutral meson fields. For this purpose we simplified and applied the complete IW method. The results obtained are in agreement with those obtained by the method of Dirac[2-4] in all cases.

II. THE METHOD OF INFELD AND WALLACE

We consider a four-dimensional space with coordinates x_μ, where x_0 is the time coordinate. Greek indices take the values 0, 1, 2, 3 and Latin indices the values 1, 2, 3. A repeated index denotes summation over its range, and we use u to denote $+(u_\mu u^\mu)^{\frac{1}{2}}$. The velocity of light is taken as unity. The metric tensor of flat space is given by

$$g_{\mu\nu}=0 \quad \text{if} \quad \mu\neq\nu, \quad g_{00}=-g_{11}=-g_{22}=-g_{33}=1. \tag{1}$$

We consider a field whose energy-momentum tensor $T_{\mu\nu}$ satisfies the equation

$$\partial T_{\mu\nu}/\partial x_\nu=0, \tag{2}$$

except at singularities in the field. If Φ_m and Φ_0 are functions satisfying the equations

$$\partial^2\Phi_m/\partial x_s\partial x^s=\partial T_{m0}/\partial x_0, \quad \partial^2\Phi_0/\partial x_s\partial x^s=\partial T_{00}/\partial x_0, \tag{3}$$

we can write for Eq. (2)

$$\begin{aligned}\partial T_{mn}/\partial x_n-\partial^2\Phi_m/\partial x_s\partial x^s=0,\\ \partial T_{0n}/\partial x_n-\partial^2\Phi_0/\partial x_s\partial x^s=0.\end{aligned} \tag{4}$$

The volume integrals of these expressions vanish if the volume does not include any field singularities; if S is the surface bounding such a volume, we obtain by Gauss's law

$$\begin{aligned}\int_S(T_{mn}-\partial\Phi_m/\partial x^n)N^n dS=0,\\ \int_S(T_{0n}-\partial\Phi_0/\partial x^n)N^n dS=0,\end{aligned} \tag{5}$$

where N^n is the nth direction cosine of the normal to dS. By a proper choice of the functions Φ_m and Φ_0, which are each arbitrary to within a solution of Laplace's equation, the integrals in (5) may be equated to zero even if one singularity is enclosed by the surface. In the following it is understood that in all surface integrals considered S does in fact enclose a singularity.

Equations (5) are the equations of motion of the simplified IW method. The method is ambiguous because it includes no rules for choosing the arbitrary harmonic functions, except for the requirement that only three of the Eqs. (5) should be independent, as this is sufficient to determine the motion.

The complete IW method is based on the linearized field equations of general relativity (which are obtained from the complete field equations by neglecting nonlinear gravitational terms). Interactions of the gravitational and nongravitational fields are also neglected; this is equivalent to using the metric tensor (1) of flat space in the calculation of $T_{\mu\nu}$. The resulting equations provide a simple way of introducing a scheme which

[2] H. J. Bhabha, Proc. Roy. Soc. (London) **A172**, 384 (1939).
[3] Harish-Chandra, Proc. Roy. Soc. (London) **A185**, 269 (1946).
[4] P. Havas, Phys. Rev. **87**, 309 (1952).
[5] J. A. McLennan, Jr., and P. Havas, Phys. Rev. **87**, 898 (1952); P. Havas, Phys. Rev. **87**, 898 (1952).
[6] M. Mathisson, Acta Phys. Polon. **6**, 167 (1937).
[7] J. Lubanski, Acta Phys. Polon. **6**, 356 (1937).
[8] Einstein, Infeld, and Hoffmann, Ann. Math. **39**, 65 (1938); A. Einstein and L. Infeld, Ann. Math. **41**, 455 (1940); A. Einstein and L. Infeld, Can. J. Math. **1**, 209 (1949).
[9] L. Infeld and P. R. Wallace, Phys. Rev. **57**, 797 (1940); hereafter referred to as IW.
[10] R. C. Majumdar and A. S. Apte, Phys. Rev. **74**, 538 (1948); hereafter referred to as MA.

includes all of the requirements of special relativity, while insuring that the special relativistic results are consistent with general relativity. The linearized equations obtained are[11]

$$\partial^2\gamma_{\mu\nu}/\partial x_\sigma\partial x^\sigma=-2T_{\mu\nu}, \tag{6}$$

where $\gamma_{\mu\nu}$ is a tensor which is related to the deviation of the metric from that for flat space, and which is subject to the coordinate conditions

$$\partial\gamma_{\mu\nu}/\partial x_\nu=0. \tag{7}$$

Combining Eqs. (6) and (7) we get

$$\frac{\partial^2\gamma_{\mu n}}{\partial x_0\partial x^0}-\frac{\partial^2\gamma_{\mu 0}}{\partial x_0\partial x^n}+2T_{\mu n}=\frac{\partial}{\partial x_\sigma}\left(\frac{\partial\gamma_{\mu n}}{\partial x^\sigma}-\frac{\partial\gamma_{\mu\sigma}}{\partial x^n}\right). \tag{8}$$

The right-hand side of Eq. (8) is the divergence of a skew-symmetric expression, taken with respect to one of the indices in which it is skew-symmetric. According to a well-known lemma[9] the surface integral of such an expression over a closed surface vanishes. Thus

$$\int_S\frac{\partial}{\partial x_\sigma}\left(\frac{\partial\gamma_{\mu n}}{\partial x^\sigma}-\frac{\partial\gamma_{\mu\sigma}}{\partial x^n}\right)N^n dS=0, \tag{9}$$

or

$$\int_S\left(\frac{\partial^2\gamma_{\mu n}}{\partial x_0\partial x^0}-\frac{\partial^2\gamma_{\mu 0}}{\partial x_0\partial x^n}+2T_{\mu n}\right)N^n dS=0. \tag{10}$$

Equations (10) are the equations of motion of the complete method of IW. In order to find their exact form for a particular field, we have to consider Eqs. (6) and (7). We shall show, however, that for the purpose of obtaining the equations of motion it is not necessary to solve these equations completely. We shall also find that the requirement that the solutions to Eq. (6) satisfy Eq. (7) and reduce to those for a mass point if the "mesic charge" of the particle is put equal to zero, determines the equations of motion unambiguously.

III. THE MESON FIELDS

We first list the meson field equations and some of their solutions; for a more detailed discussion, see reference 4 and the references listed there.

The field equations are[12]

$$F_\mu=-\partial U/\partial x^\mu, \qquad -\partial F_\mu/\partial x_\mu+\chi^2U=4\pi\rho, \tag{11s}$$

$$G_{\mu\nu}=\partial U_\nu/\partial x^\mu-\partial U_\mu/\partial x^\nu,\ \partial G_{\mu\nu}/\partial x_\mu+\chi^2U_\nu=4\pi j_\nu. \tag{11v}$$

Here F_μ and $G_{\mu\nu}$ are the field strength components for the scalar and the vector field with potentials U and U_ν, respectively. χ is a parameter, which in quantized theory is related to the rest mass μ of the mesons associated with the field by $\chi=\mu c/\hbar$. ρ and j_ν are the charge density and four-current density of mesic charge.

[11] See, e.g., P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Inc., New York, 1942), Chapter XII.

[12] In the following, corresponding equations for the scalar and vector fields are written with the same number, but with the suffix s and v, respectively.

In this paper, we are concerned only with simple poles in the meson fields. For a particle which is a simple pole with mesic charge g, and whose coordinates and four-velocity are z_μ and v_μ, respectively, ρ and j_ν are given by

$$\rho=g\int_{-\infty}^{\infty}\delta(x_0-z_0)\delta(x_1-z_1)\delta(x_2-z_2)\delta(x_3-z_3)d\tau, \tag{12s}$$

$$j_\nu=g\int_{-\infty}^{\infty}v_\nu\delta(x_0-z_0)\delta(x_1-z_1)\delta(x_2-z_2)\delta(x_3-z_3)d\tau, \tag{12v}$$

where δ is Dirac's delta function, and τ the proper time of the particle.

Particular solutions to the field equations (11) are given by

$${}_{\rm ret}U=g\frac{1}{\kappa({}_r\tau)}-g\chi\int_{-\infty}^{{}_r\tau}\frac{1}{s}J_1(\chi s)d\tau, \tag{13s}$$

$${}_{\rm sym}U=\tfrac{1}{2}g\left(\frac{1}{\kappa({}_r\tau)}-\frac{1}{\kappa({}_a\tau)}\right)-\tfrac{1}{2}g\chi\int_{-\infty}^{\infty}\frac{1}{s}J_1(\chi s)d\tau+\tfrac{1}{2}g\chi\int_{{}_r\tau}^{{}_a\tau}\frac{1}{s}J_1(\chi s)d\tau, \tag{14s}$$

$${}_{\rm ret}U_\nu=g\frac{v_\nu({}_r\tau)}{\kappa({}_r\tau)}-g\chi\int_{-\infty}^{{}_r\tau}\frac{v_\nu}{s}J_1(\chi s)d\tau, \tag{13v}$$

$${}_{\rm sym}U_\nu=\tfrac{1}{2}g\left(\frac{v_\nu({}_r\tau)}{\kappa({}_r\tau)}-\frac{v_\nu({}_a\tau)}{\kappa({}_a\tau)}\right)-\tfrac{1}{2}g\chi\int_{-\infty}^{\infty}\frac{v_\nu}{s}J_1(\chi s)d\tau+\tfrac{1}{2}g\chi\int_{{}_r\tau}^{{}_a\tau}\frac{v_\nu}{s}J_1(\chi s)d\tau, \tag{14v}$$

where $s_\mu=x_\mu-z_\mu(\tau)$ and $\kappa=s_\mu v^\mu$. ${}_a\tau$ and ${}_r\tau$ are the proper times at the advanced and retarded points, respectively. These are given by $s_\mu s^\mu=0$ with $s_0<0$ and >0, respectively. J_1 is the Bessel function of order 1. The retarded and symmetric solutions (13) and (14), which are analogous to the corresponding solutions to Maxwell's equations, differ by the radiation potential, which is finite at the particle, together with all its derivatives. The first term of (13) or (14) is a solution to (11) with $\chi=0$; Eqs. (11v) then reduce to the electromagnetic field equations. The terms in (13) and (14) which are independent of χ may be written:

$${}_{\rm sym}U^0=g\sum_{l=1}^{\infty}\frac{1}{(2l-2)!}\frac{d^{2l-2}}{d\tau^{2l-2}}(r^{2l-3}), \tag{15s}$$

$${}_{\rm ret}U^0={}_{\rm sym}U^0-g\sum_{l=1}^{\infty}\frac{1}{(2l-1)!}\frac{d^{2l-1}}{d\tau^{2l-1}}(r^{2l-2}), \tag{16s}$$

$${}_{\rm sym}U_\mu{}^0=g\sum_{l=1}^{\infty}\frac{1}{(2l-2)!}\frac{d^{2l-2}}{d\tau^{2l-2}}(v_\mu r^{2l-3}), \tag{15v}$$

$${}_{\rm ret}U_\mu{}^0={}_{\rm sym}U_\mu{}^0-g\sum_{l=1}^{\infty}\frac{1}{(2l-1)!}\frac{d^{2l-1}}{d\tau^{2l-1}}(v_\mu r^{2l-2}), \tag{16v}$$

where $r=(-s_\lambda s^\lambda)^{\frac{1}{2}}$. The expressions (15) and (16) may be obtained in several ways; one of these is described in the appendix, and another has been suggested by Page and Adams.[13]

When the field strengths are calculated from (13) and (14) by differentiation, the parts (15) and (16) give rise to terms independent of χ. The integrals contribute terms of two kinds. Those arising from the differentiation at the limits, which are discontinuous at the particle, cannot contribute to the equations of motion, as will be shown in Sec. V. The remaining terms, obtained by differentiation under the integral sign, are finite and continuous at the particle and thus behave just like an external field. The total field is given by

$$ {}_{\rm sym}F_\mu = {}_{\rm sym}F_\mu{}^o + {}_{\rm ext}F_\mu + \tfrac{1}{2}g\chi^2\left(\frac{s_\mu(_r\tau)}{\kappa(_r\tau)} - \frac{s_\mu(_a\tau)}{\kappa(_a\tau)}\right) + \tfrac{1}{2}g\chi^2\int_{_r\tau}^{^a\tau}\frac{s_\mu J_2(\chi s)}{s^2}d\tau - \tfrac{1}{2}g\chi^2\int_{-\infty}^{^r\tau}\frac{s_\mu J_2(\chi s)}{s^2}d\tau, \tag{17s}$$

$$ {}_{\rm ret}F_\mu = {}_{\rm ret}F_\mu{}^o + {}_{\rm ext}F_\mu + \tfrac{1}{2}g\chi^2\frac{s_\mu(_r\tau)}{\kappa(_r\tau)} - g\chi^2\int_{-\infty}^{^r\tau}\frac{s_\mu J_2(\chi s)}{s^2}d\tau, \tag{18s}$$

$$ {}_{\rm sym}G_{\mu\nu} = {}_{\rm sym}G_{\mu\nu}{}^o + {}_{\rm ext}G_{\mu\nu} + \tfrac{1}{4}g\chi^2\left(\frac{s_\mu(_r\tau)v_\nu(_r\tau)}{\kappa(_r\tau)} - \frac{s_\nu(_r\tau)v_\mu(_r\tau)}{\kappa(_r\tau)}\right) - \tfrac{1}{4}g\chi^2\left(\frac{s_\mu(_a\tau)v_\nu(_a\tau)}{\kappa(_a\tau)} - \frac{s_\nu(_a\tau)v_\mu(_a\tau)}{\kappa(_a\tau)}\right) $$
$$ + \tfrac{1}{2}g\chi^2\int_{_r\tau}^{^a\tau}\frac{s_\mu v_\nu - s_\nu v_\mu}{s^2}J_2(\chi s)d\tau - \tfrac{1}{2}g\chi^2\int_{-\infty}^{^r\tau}\frac{s_\mu v_\nu - s_\nu v_\mu}{s^2}J_2(\chi s)d\tau, \tag{17v}$$

$$ {}_{\rm ret}G_{\mu\nu} = {}_{\rm ret}G_{\mu\nu}{}^o + {}_{\rm ext}G_{\mu\nu} + \tfrac{1}{2}g\chi^2\left(\frac{s_\mu(_r\tau)v_\nu(_r\tau)}{\kappa(_r\tau)} - \frac{s_\nu(_r\tau)v_\mu(_r\tau)}{\kappa(_r\tau)}\right) - g\chi^2\int_{-\infty}^{^r\tau}\frac{s_\mu v_\nu - s_\nu v_\mu}{s^2}J_2(\chi s)d\tau, \tag{18v}$$

where the superscript o denotes that part of the fields that arises from (15) or (16), and J_2 is the Bessel function of order 2. The expressions (17) and (18) may then be used to calculate the energy-momentum tensors, which are given by

$$4\pi T_{\mu\nu} = F_\mu F_\nu - \tfrac{1}{2}g_{\mu\nu}(F_\sigma F^\sigma - \chi^2 U^2), \tag{19s}$$

$$4\pi T_{\mu\nu} = G_{\mu\sigma}G^\sigma{}_\nu + \tfrac{1}{4}g_{\mu\nu}G_{\rho\sigma}G^{\rho\sigma} + \chi^2 U_\mu U_\nu - \tfrac{1}{2}\chi^2 g_{\mu\nu}U_\sigma U^\sigma. \tag{19v}$$

IV. THE SIMPLIFIED METHOD OF INFELD AND WALLACE AND THE CALCULATIONS OF MAJUMDAR AND APTE

The simplified method of IW[9] has been applied to the neutral meson fields by MA.[10] By choosing the harmonic functions in (5) to be

$$\Phi_m = m_0 v_m/r, \quad \Phi_0 = m_0/r, \tag{20}$$

in both the vector and scalar cases, they obtain the equations of motion

$$d(m_0 v_\mu)/d\tau = g\bar{F}^\nu(g_{\mu\nu} - v_\mu v_\nu), \tag{21s}$$

$$d(m_0 v_\mu)/d\tau = g\bar{G}_{\mu\nu}v^\nu, \tag{21v}$$

where $\bar{F}_\nu$ and $\bar{G}_{\mu\nu}$ are those parts of the fields (17) or (18) which are finite and continuous at the particle; they are evaluated at the position of the singularity. In the retarded case, Eq. (21v) is identical with that obtained by Bhabha,[2] but Eq. (21s) does not agree with that obtained by Harish-Chandra[3] applying Dirac's method,

$$d[(m_0 - g\bar{U})v_\mu]/d\tau = g\bar{F}_\mu. \tag{22s}$$

Similar remarks apply to the symmetric case.[4]

Although Dirac's method does not lead unambiguously to Eq. (22s), it does require that the equations of motion have the form

$$dP_\mu/d\tau = g\bar{F}_\mu, \tag{23s}$$

where P_μ is some function of τ, and is thus incompatible with the result (21s).[14]

On the other hand, Eq. (22s) is compatible with the simplified IW method; it follows from Eqs. (5) by the choice

$$\Phi_m = (m_0 - g\bar{U})v_m/r, \quad \Phi_0 = (m_0 - g\bar{U})/r, \tag{24s}$$

instead of (20) for the arbitrary harmonic functions.

V. THE COMPLETE METHOD OF INFELD AND WALLACE FOR MESON FIELDS

The explicit determination of the equations of motion from Eq. (10) of the complete Infeld-Wallace method can be considerably simplified by an investigation of the way in which various terms can contribute to the results. For this discussion it is convenient to use the "special coordinate system" in which the particle is instantaneously at rest at the origin ($z_k = v_k = 0$, $v_0 = 1$) and to take the surface of integration in (10) to be a sphere with the particle at its center.

The parts of the fields given by Eqs. (15) or (16) can be expressed as sums of terms of the form

$$\frac{x_{k_1}x_{k_2}x_{k_3}\cdots x_{k_p}}{r^q}f(\tau). \tag{25}$$

[13] L. Page and N. I. Adams, Jr., *Electrodynamics* (D. Van Nostrand Company, Inc., New York, 1940), Chapter IV.

[14] MA state that Eq. (21s) is "practically the same" as Eq. (22s), "provided we absorb the term proportional to acceleration in the definition of the inertial mass." This, however, is not possible, as then the rest mass would no longer be constant.

The remaining terms in the field quantities can be expressed in the same form by using their Taylor expansions about the location of the particle, as they are finite at the origin. As the energy-momentum tensors are quadratic in the field quantities, they also may be written as sums of terms of the form (25).

We say that a term like (25) behaves like r^{p-q}; in order to contribute an expression to the integrals in (10), taken over a sphere, which is independent of the surface of integration, such a term must behave like r^{-2}. In addition, a term which before multiplication by the direction cosine $N^n=x^n/r$ has p even will vanish by symmetry upon integration.

As noted before, the radiation field is finite at the particle, and thus the singularities of expression (16) are the same as those of (15). For each undifferentiated term in (15) we have $p-q=2l-3$ and q odd. Differentiation of an expression of the form (25) with respect to the proper time either leaves p and q unchanged and increases the order of a time differentiation, or decreases $p-q$ by one and introduces a factor v_k, q changing by two. To obtain a nonvanishing expression for the field strengths in the special coordinate system the potentials can contain at most one factor v_k; therefore we have to remove all but at most one v_k by differentiation. Thus the $2l-2$ differentiations of a term of (15) can decrease $p-q$ by at most $l-1$ ($l\geq 1$) and the resultant $p-q$ will equal at least $l-2$. Thus the most singular term of the potentials has $p-q=-1$ and arises from the first two terms of the series in (15). Therefore the terms in U^2 of the scalar case (19s) could only contribute through products of such terms. For these products, however, p equals zero and thus the integral vanishes.

Differentiation with respect to the space coordinates of the field point decreases $p-q$ by one and time differentiation either leaves $p-q$ unchanged or introduces a factor v_k, q remaining odd. Therefore the most singular part of the field strengths has $p-q=-2$ (with p and q odd). The only terms in $T_{\mu\nu}$ which can contribute directly to the equations of motion are products of this part and the finite, continuous parts of the field strengths $\bar{F}_\mu$ or $\bar{G}_{\mu\nu}$, evaluated at the position of the particle. Its products with the finite, discontinuous parts of the field strengths vanish on integration, as these parts have $p-q$ at least zero and p can be seen to be odd. Similarly all other terms in $T_{\mu\nu}$ with $p-q=-2$ can easily be seen to have even p and thus vanish on integration. Thus we have

$$\begin{aligned}
&4\pi T_{mn}\sim-(g/r^3)(x_m\bar{F}_n+x_n\bar{F}_m-g_{mn}x_s\bar{F}^s),\\
&4\pi T_{0n}\sim-(g/r^3)(x_n\bar{F}_0+x_kv^k\bar{F}_n),\qquad(26s)\\
&4\pi T_{00}\sim(g/r^3)(x_k\bar{F}^k+x_kv^k\bar{F}_0),\\
&4\pi T_{mn}\sim-(g/r^3)(x_m\bar{G}_{n0}+x_n\bar{G}_{m0}-g_{mn}x^s\bar{G}_{s0}),\\
&4\pi T_{0n}\sim(g/r^3)(x_kv^k\bar{G}_{n0}),\qquad(26v)\\
&4\pi T_{00}\sim(g/r^3)(x^k\bar{G}_{k0}),
\end{aligned}$$

where we use $\sim$ to mean that we have omitted some terms which cannot contribute to the integration. However, we have kept some terms which vanish at the origin, as they may be needed in considering the coordinate conditions.

We next consider the way in which the $T_{\mu\nu}$ can contribute to the equations of motion via the $\gamma_{\mu\nu}$. We shall write

$$\gamma_{\mu\nu}=\bar{\gamma}_{\mu\nu}+\Gamma_{\mu\nu},\qquad(27)$$

where we use $\bar{\gamma}_{\mu\nu}$ for the solution of the homogeneous equations corresponding to Eq. (6) and $\Gamma_{\mu\nu}$ for a particular solution of (6). Applying the method of the appendix to Eq. (6) we find that the $\Gamma_{\mu\nu}$ behave like

$$r^2T_{\mu\nu}+d^2(r^4T_{\mu\nu})/d\tau^2+\cdots.\qquad(28)$$

It then follows from the previous considerations that the $\Gamma_{\mu\nu}$ do not contain any singularities which could contribute to the equations of motion.

The solutions to the homogeneous equation corresponding to Eq. (6) are

$$4\pi_{\text{asym}}\bar{\gamma}_{\mu\nu}=\sum_{l=1}^{\infty}\frac{2}{(2l-2)!}\frac{d^{2l-2}}{d\tau^{2l-2}}(A_{\mu\nu}r^{2l-3}),\qquad(29)$$

$$4\pi_{\text{ret}}\bar{\gamma}_{\mu\nu}=4\pi_{\text{asym}}\bar{\gamma}_{\mu\nu}-\sum_{l=1}^{\infty}\frac{2}{(2l-1)!}\frac{d^{2l-1}}{d\tau^{2l-1}}(A_{\mu\nu}r^{2l-2}).\qquad(30)$$

Here $A_{\mu\nu}$ is a symmetric tensor, which is a function of the proper time only. We take the retarded solution if we use retarded expressions for the meson fields, and correspondingly for the symmetric field. The solutions (29) and (30) correspond to a simple pole in the gravitational field, i.e., a mass point. The only terms in (29) or (30) which can contribute to the integration and their second derivatives required in (10) are

$$\begin{aligned}
&4\pi\bar{\gamma}_{\mu\nu}\sim 2A_{\mu\nu}/r,\\
&4\pi\partial^2\bar{\gamma}_{mn}/\partial x_0\partial x^0\sim-2A_{mn}x_k\dot{v}^k/r^3,\\
&4\pi\partial^2\bar{\gamma}_{0n}/\partial x_0\partial x^0\sim-2A_{0n}x_k\dot{v}^k/r^3,\qquad(31)\\
&4\pi\partial^2\bar{\gamma}_{m0}/\partial x_0\partial x^n\sim 2\dot{A}_{m0}x_n/r^3,\\
&4\pi\partial^2\bar{\gamma}_{00}/\partial x_0\partial x^n\sim 2\dot{A}_{00}x_n/r^3,
\end{aligned}$$

where the dots denote differentiation with respect to the proper time. All other terms will give terms in the integrand which either have $q-p\neq 2$, or which contain a factor v_k, and hence vanish in the special coordinate system.

Using $\int_s r^{-3}x_kN_n\,dS=-(4\pi/3)g_{kn}$, we obtain for the equations of motion in the special coordinate system, using Eqs. (10), (26), and (31),

$$-\tfrac{4}{3}A_{\mu n}\dot{v}^n-\dot{A}_{\mu 0}+g\bar{F}_\mu=0,\qquad(32s)$$

$$-\tfrac{4}{3}A_{\mu n}\dot{v}^n-\dot{A}_{\mu 0}+g\bar{G}_{\mu 0}=0.\qquad(32v)$$

These equations, together with the requirements that the $\gamma_{\mu\nu}$ must satisfy the coordinate conditions (7) and

must reduce to those for a mass point when $g=0$, determine the $A_{\mu\nu}$ completely.

For the coordinate conditions, we consider the equation

$$\frac{\partial^2}{\partial x_\mu \partial x^\rho}\left(\frac{\partial \gamma_{\mu\nu}}{\partial x_\nu}\right)=-2\frac{\partial T_{\mu\nu}}{\partial x_\nu}=0, \tag{33}$$

which follows from the field equations (6) and the conservation law (2), except at the singularity. The $\partial\gamma_{\mu\nu}/\partial x_\nu$ are thus solutions to the homogeneous wave equation except at the singularity. Finite terms in these expressions may be represented by their generalized Fourier expansions in terms of solutions to the homogeneous equation, and hence can be subtracted out by suitable choice of such solutions. This will not affect the equations of motion since such solutions correspond to finite terms in the $\gamma_{\mu\nu}$ which do not influence the equations of motion. By the method of the appendix, singular solutions to (33) are of the form (25) with q odd, since they are obtained by differentiation of $1/r$. But by our discussion of the manner in which a differentiation can affect such terms, they must correspond to terms in $T_{\mu\nu}$ with q odd. These must have $p-q=-2$ to contribute singular terms to (33), and thus are just the terms given in (26).

The terms in $\Gamma_{\mu\nu}$ which can contribute to the coordinate conditions can again be obtained by the method of the appendix. We get

$$\begin{aligned}
&4\pi\Gamma_{mn}\sim(g/r)(x_m\bar{F}_n+x_n\bar{F}_m-g_{mn}x_k\bar{F}^k),\\
&4\pi\Gamma_{0n}\sim(g/r)(x_n\bar{F}_0+x_kv^k\bar{F}_n),\\
&4\pi\Gamma_{00}\sim-(g/r)(x_k\bar{F}^k+x_kv^k\bar{F}_0),
\end{aligned} \tag{34s}$$

$$\begin{aligned}
&4\pi\Gamma_{mn}\sim(g/r)(x_m\bar{G}_{n0}+x_n\bar{G}_{m0}-g_{mn}x^k\bar{G}_{k0}),\\
&4\pi\Gamma_{0n}\sim(g/r)(x_kv^k\bar{G}_{n0}),\\
&4\pi\Gamma_{00}\sim-(g/r)(x^k\bar{G}_{k0}).
\end{aligned} \tag{34v}$$

These are the solutions to the lowest approximation only; the higher approximations will give either terms with $q-p<1$, or terms with at least a factor v_k in the coordinate conditions. Omitting terms after differentiation of the $\gamma_{\mu\nu}$ and the $\Gamma_{\mu\nu}$ which are nonsingular, or which contain a factor v_k, we obtain for the coordinate conditions in the special coordinate system

$$-A_{\mu n}x^n/r^3-\dot{A}_{\mu 0}/r+g\bar{F}_\mu/r=0, \tag{36s}$$

$$-A_{\mu n}x^n/r^3-\dot{A}_{\mu 0}/r+g\bar{G}_{\mu 0}/r=0. \tag{36v}$$

Combining Eqs. (36) and (32), we obtain

$$A_{\mu n}\left(\frac{x^n}{r^3}-\frac{\dot{v}^n}{3r}\right)=0, \tag{37}$$

and hence all the $A_{\mu\nu}$ except A_{00} vanish in the special coordinate system. Thus $A_{\mu\nu}$ must be of the form $Mv_\mu v_\nu$, where M is a scalar function of τ; in order that the solutions to the homogeneous gravitational equation reduce to those for a mass point when $g=0$, M must reduce to the rest mass m_0 in that case.

In terms of M the equations of motion (32) become

$$d(Mv_m)/d\tau=g\bar{F}_m,\quad dM/d\tau=g\bar{F}_0, \tag{38s}$$

$$d(Mv_m)/d\tau=g\bar{G}_{m0},\quad dM/d\tau=0. \tag{38v}$$

The fourth equation in each case determines M to within a constant; the value of this constant is determined by the previous requirement that M reduce to m_0 when $g=0$. Thus we obtain

$$M=m_0-g\bar{U}, \tag{39s}$$

$$M=m_0, \tag{39v}$$

without any ambiguity. The requirements of the complete IW method thus determine completely the equations of motion of a point particle which is a simple pole in both the mesic and the gravitational field; the appearance of the inertial terms in these equations is due to the equivalence of gravitational and inertial mass.

Inserting the expressions (39) into Eqs. (38), using the fields given in Sec. III and dropping the restriction to the special coordinate system, we finally obtain the following equations of motion:

Retarded field case:

$$m_0\dot{v}^\mu-\tfrac{1}{3}g^2(\ddot{v}^\mu+v^\mu\dot{v}^2)-\tfrac{1}{2}g^2\chi^2v^\mu+g^2\chi^2\int_{-\infty}^{\tau}\frac{s^\mu}{s^2}J_2(\chi s)d\tau+g^2\chi\frac{d}{d\tau}\left[v^\mu\int_{-\infty}^{\tau}\frac{1}{s}J_1(\chi s)d\tau\right]=g_{\mathrm{ext}}F^\mu+g\frac{d}{d\tau}({}_{\mathrm{ext}}Uv^\mu), \tag{40s}$$

$$m_0\dot{v}^\mu-\tfrac{2}{3}g^2(\ddot{v}^\mu+v^\mu\dot{v}^2)-g^2\chi^2v_\nu\int_{-\infty}^{\tau}\frac{s^\mu v^\nu-s^\nu v^\mu}{s^2}J_2(\chi s)d\tau=g_{\mathrm{ext}}G^\mu{}_\nu v^\nu; \tag{40v}$$

Symmetric field case:

$$m_0\dot{v}^\mu+\tfrac{1}{2}g^2\chi^2\int_{-\infty}^{\infty}\frac{s^\mu}{s^2}J_2(\chi s)d\tau+\tfrac{1}{2}g^2\chi\frac{d}{d\tau}\left[v^\mu\int_{-\infty}^{\infty}\frac{1}{s}J_1(\chi s)d\tau\right]=g_{\mathrm{ext}}F^\mu+g\frac{d}{d\tau}({}_{\mathrm{ext}}Uv^\mu), \tag{41s}$$

$$m_0\dot{v}^\mu-\tfrac{1}{2}g^2\chi^2v_\nu\int_{-\infty}^{\infty}\frac{s^\mu v^\nu-s^\nu v^\mu}{s^2}J_2(\chi s)d\tau=g_{\mathrm{ext}}G^\mu{}_\nu v^\nu. \tag{41v}$$

VI. DISCUSSION

Our equations of motion (40) and (41), obtained by the use of the complete IW method, agree with those obtained previously by Dirac's method.[2-4] As in the case of the electromagnetic field treated by IW (which are included in our calculations as the special case $\chi=0$ of the vector meson field), no approximations were used beyond those of the linearized equations of general relativity and the neglect of gravitational-mesic interactions. It is this use of general relativity which allows us to avoid the ambiguity of other methods. In contrast to Dirac's method, we also avoid the necessity of subtracting out infinite self-energy terms explicitly and of evaluating surface integrals in four dimensions; the evaluations of such integrals in three-dimensional space has been considerably simplified by extending the symmetry arguments of IW.

The method used here can be extended without difficulty to the case of a particle which is a singularity of several noninteracting fields. The energy-momentum tensor for noninteracting fields is simply the sum of those for the separate fields. Similarly $\Gamma_{\mu\nu}$ is the sum of the $\Gamma_{\mu\nu}$'s for the separate fields, because the discussion of the coordinate condition can be carried through as before. As M must still reduce to m_0 if the constants describing the coupling between the particle and the different fields are set equal to zero, the terms which must be added to m_0 to get M for a combination of noninteracting fields are just the sum of the terms required for the separate fields. Thus, the rate of change of $m_0 v_\mu$ for such a combination of fields is simply the sum of the rates of change due to the separate fields. Therefore the equations of motion of particles interacting with a combination of scalar and vector meson fields and of electromagnetic fields can be written down without further calculations. The resulting equations again agree with those obtained by Dirac's method.[15]

The authors are indebted to Dr. B. Kaufman for her valuable criticism of the manuscript.

APPENDIX

We discuss the solution of the inhomogeneous wave equation,

$$\partial^2\phi/\partial x_\sigma \partial x^\sigma = -\rho, \tag{A1}$$

by a method due to Infeld.[16] ϕ and ρ may be scalars or components of vectors or tensors. We assume a solution of the form

$$\phi=\sum_{k=0}^{\infty}\phi_k\lambda^k, \tag{A2}$$

where the parameter λ is introduced by using the auxiliary time $w=\lambda x_0$. Substituting in (A1) one obtains by equating coefficients of λ^k

$$\partial^2\phi_0/\partial x_\sigma\partial x^\sigma=-\rho, \quad \partial^2\phi_k/\partial x_\sigma\partial x^\sigma=\partial^2\phi_{k-2}/\partial w^2. \tag{A3}$$

Assuming a solution of the form $\phi_k=d^k f_k/dw^k$ (k not summed), and putting $\lambda=1$, we obtain

$$\partial^2 f_0/\partial x_\sigma\partial x^\sigma=-\rho, \quad \partial^2 f_k/\partial x_\sigma\partial x^\sigma=f_{k-2}. \tag{A4}$$

Thus the inhomogeneous wave equation may be replaced by a set of Poisson equations.

If the density function can be expressed as a sum of terms like (25) in the coordinates of the field point with respect to some origin (such as the location of the particle, as in the case of the $T_{\mu\nu}$ in the gravitational field equations), the solutions behave like

$$r^2\rho+\frac{d^2}{dr^2}(r^4\rho)+\cdots. \tag{A5}$$

If we take $\rho=0$ except at a single point, and $A(\tau)/r$ as the solution ϕ_0 to the lowest approximation, we easily obtain the solution

$${}_{\rm sym}\phi=\sum_{l=1}^{\infty}\frac{1}{(2l-2)!}\frac{d^{2l-2}}{d\tau^{2l-2}}(A\tau^{2l-3}). \tag{A6}$$

This expression is symmetric with respect to past and future. One needs only even-order derivatives, as (A4) relates only terms differing in order by two. We have replaced the derivatives with respect to x_0 by derivatives with respect to the proper time of the singularity; these are written with ordinary d's to remind us that we must consider the implicit dependence of the differentiated quantities on τ due to the motion of the singularity. With respect to the coordinates of the field point the differentiation is partial, however.

Another solution to the inhomogeneous wave equation may be obtained by including odd-order derivatives by taking $\phi_1=-dA/dw$. We then obtain the retarded solution given by

$${}_{\rm ret}\phi={}_{\rm sym}\phi-\sum_{l=1}^{\infty}\frac{1}{(2l-1)!}\frac{d^{2l-1}}{d\tau^{2l-1}}(A\tau^{2l-2}). \tag{A7}$$

[15] Compare P. Havas, Phys. Rev. **94**, 882 (1954).
[16] L. Infeld, Phys. Rev. **53**, 836 (1938).

Application of Variational Methods to Intermediate and High-Energy Scattering*

E. GERJUOY, *University of Pittsburgh, Pittsburgh, Pennsylvania*

AND

DAVID S. SAXON,† *National Bureau of Standards, Los Angeles, California*

(Received December 9, 1953)

The application of Schwinger's variational principle to scattering by a Yukawa potential is studied. A detailed comparison of the Born approximation, variational, and exact results is given in the energy region from 20 to 150 Mev for a potential strength and range appropriate to the ^3S neutron-proton interaction. Comparison with the second Born approximation is also made. The utility of the variational principle is discussed and consideration is given to the high-energy limit. It is judged that for most potentials, the variational formulation without elaborate trial functions will be better than the Born approximation.

The relation between the variational principles for the total scattering amplitude and for the phase shifts is analyzed. A new formulation intermediate between these two is presented as a compromise between the simplicity of the former and the accuracy of the latter. With the potential studied, use of this formulation leads to considerably better results for the total amplitude than does the original principle.

I. INTRODUCTION

IN the simplest and most fundamental quantum-mechanical scattering problem, the scattering of a particle by a fixed potential, which also yields the scattering of one particle by another in the center-of-mass frame, there exists an important energy region where no entirely suitable methods are available for the calculation of differential cross sections. We refer of course to those intermediate energies for which expansion in partial waves requires the determination of numerous phase shifts, and for which on the other hand, such high-energy approximations as the Born approximation are inadequate for estimating these phase shifts and, therefore, for estimating unexpanded or closed form approximations to the differential cross section. Consequently in this intermediate energy range the computations are tedious even for the simplest potentials, involving as they do numerical integration of the radial differential equation for each of the numerous partial waves. Moreover the resultant cross sections vary with energy and other parameters in so complicated a fashion, and in a manner so tenuously related to the detailed shape of the potential, that such calculations are actually not very enlightening qualitatively. The result is that the admittedly inadequate Born approximation, which has the virtue of simplicity, is frequently used in the hope that it can serve as a not entirely incorrect qualitative guide.

For the case of neutron-proton scattering this intermediate energy region begins at a laboratory energy E_L of about 20 Mev, where the contributions of higher (than S) partial waves become significant.[1] With a Yukawa well the validity of the Born approximation has seemed questionable at energies as high as 250 Mev "where the potential picture is getting to be quite far-fetched," since the second Born approximation is significantly different from the first, even up to these energies.[2] In the present paper we investigate the utility of Schwinger's variational expression for the total scattering amplitude[3] in bridging this gap for central field neutron-proton scattering with Yukawa wells, including exchange forces. The calculations have been performed for a potential $(V_0a/r)e^{-r/a}$ such that $V_0=53.8$ Mev and $a=1.35\times10^{-13}$ cm, corresponding to the ^3S neutron-proton interaction. In subsequent sections we describe in detail the comparison of the variational results for several varieties of trial functions with the "exact" differential cross sections obtained by numerical integration of Schrödinger's equation,[4] and with the results in Born approximation.[5] Briefly we may state that the Born approximation turns out to be quite good, particularly in the absence of exchange forces. As described in detail below, with an improved formulation and appropriately chosen wave function, the variational method is superior to Born approximation, rather more so when exchange forces are included. Naturally the variational results are achieved at the expense of considerably increased computational effort. It is our belief, to be elaborated in a later section, that the Yukawa case is especially and accidentally favorable for the Born approximation, while simultaneously unfavorable for the variational principle. We feel therefore, although we are in no position to guarantee, that for other well shapes the Schwinger formulation usually will prove superior to Born approximation with relatively less complicated machinery and trial functions than those we have found necessary,[6] as confirmed

* The preparation of this paper was sponsored (in part) by the U. S. Office of Naval Research.

† Present address, University of California, Los Angeles, California.

[1] G. Snow, Phys. Rev. 87, 21 (1952).

[2] R. Jost and A. Pais, Phys. Rev. 82, 840 (1951).

[3] J. Schwinger, "Lectures on Nuclear Physics," Harvard University, 1947 (unpublished).

[4] Futterman, Osborne, and Saxon (to be published).

[5] By Born approximation we mean always first Born approximation, unless otherwise stated.

[6] It must be remarked however that our choice of well shape, Yukawa, was motivated not merely by its obvious theoretical interest, but also by the fact that for the Yukawa well the integrals

by previously reported calculations of the total cross sections for a square well.[7]

Irrespective of the degree of generality of our particular numerical results, our method of application of Schwinger's variational principle has independent theoretical interest. As originally introduced, this variational principle was written in two forms. In one, the usual expansion in spherical harmonics was made, leading to an infinite set of independent integral equations and thence to a corresponding sets of variational expressions for the phase shifts. In the other, the entire scattering amplitude was expressed in stationary form, thus giving the scattering cross section directly, without necessitating summing over the individual phase shifts. These two modes of expressing the variational principle are formally equivalent, of course. Regarded as the basis for approximate calculations, however, they have quite different attributes: The expansion is intrinsically more accurate, whereas a direct estimate of the entire scattering amplitude is intrinsically simpler. Our method has been to develop a formulation intermediate between Schwinger's two alternatives by expanding in a finite set of "partial waves" using the eigenfunctions of finite rotation and reflection operators. This, we believe, represents a useful compromise between the requirements of accuracy and simplicity.

These introductory remarks are amplified in the following sections.

II. THE VARIATIONAL PRINCIPLES

In the coordinate system in which the center of mass is at rest, Schrödinger's equation for two particles interacting according to the central potential $V(r)$ can be written in the dimensionless form

$$\nabla^2\psi(\mathbf{x})+[k^2+u(x)]\psi(\mathbf{x})=0, \tag{1}$$

where $\mathbf{x}=\mathbf{r}/a$, with a being a characteristic length (say the range), associated with the potential, and where

$$k^2=(2mE/\hbar^2)a^2, \quad u(x)=-(2m/\hbar^2)a^2V(ax), \tag{2}$$

with m the reduced mass, and E the energy in the center-of-mass system. Derivations of the variational formulation can be found, both in terms of the phase shifts[8] and of the entire scattering amplitude.[9] For convenience in describing our method, we very briefly recapitulate these derivations.

We first obtain a stationary expression for the amplitude $A(\mathbf{k}\to\mathbf{k}')$ of a wave scattered in the direction $\mathbf{k}'$, when the incident wave is in the direction $\mathbf{k}$. For this purpose Eq. (1) is re-expressed as an integral equation incorporating the proper boundary conditions, namely,

$$\psi_{\mathbf{k}}(\mathbf{x})=e^{i\mathbf{k}\cdot\mathbf{x}}+\int d\mathbf{x}'G(\mathbf{x},\mathbf{x}')u(x')\psi_{\mathbf{k}}(\mathbf{x}'), \tag{3}$$

where

$$G(\mathbf{x},\mathbf{x}')=e^{i|\mathbf{x}-\mathbf{x}'|}/4\pi|\mathbf{x}-\mathbf{x}'|, \tag{4}$$

and where the subscript $\mathbf{k}$ on ψ denotes the incident wave is in the direction $\mathbf{k}$. Examination of the asymptotic form of $\psi_{\mathbf{k}}$ for Eq. (3) then yields the rigorous expression for the amplitude of the scattered wave in the direction $\mathbf{k}'$,

$$A(\mathbf{k}\to\mathbf{k}')=\frac{1}{4\pi}\int d\mathbf{x}e^{-i\mathbf{k}'\cdot\mathbf{x}}u(x)\psi_{\mathbf{k}}(\mathbf{x})\equiv B(\mathbf{k}\to\mathbf{k}'). \tag{5}$$

From Eq. (5) the amplitude $A(-\mathbf{k}'\to-\mathbf{k})$ of the time reversed solution is

$$A(-\mathbf{k}'\to-\mathbf{k})=\frac{1}{4\pi}\int d\mathbf{x}e^{i\mathbf{k}\cdot\mathbf{x}}u(x)\psi_{-\mathbf{k}'}(\mathbf{x}) \equiv B(-\mathbf{k}'\to-\mathbf{k}). \tag{6}$$

Moreover, from Eq. (3) and the corresponding time reversed integral equation it is shown that

$$A(\mathbf{k}\to\mathbf{k}')=A(-\mathbf{k}'\to-\mathbf{k})=\frac{1}{4\pi}\int d\mathbf{x}\psi_{\mathbf{k}}(\mathbf{x})u(x)\psi_{-\mathbf{k}'}(\mathbf{x}) -\frac{1}{4\pi}\int\int d\mathbf{x}d\mathbf{x}'\psi_{\mathbf{k}}(\mathbf{x})u(x)G(\mathbf{x},\mathbf{x}')u(x')\psi_{-\mathbf{k}'}(\mathbf{x}') \equiv C(\mathbf{k}\to\mathbf{k}'). \tag{7}$$

Thereby it follows that

$$A(\mathbf{k}\to\mathbf{k}') =\frac{B(\mathbf{k}\to\mathbf{k}')B(-\mathbf{k}'\to-\mathbf{k})}{C(\mathbf{k}\to\mathbf{k}')} =\frac{1}{4\pi}\frac{\left[\int d\mathbf{x}e^{-i\mathbf{k}'\cdot\mathbf{x}}u\psi_{\mathbf{k}}\right]\left[\int d\mathbf{x}e^{i\mathbf{k}\cdot\mathbf{x}}u\psi_{-\mathbf{k}'}\right]}{\int d\mathbf{x}\psi_{\mathbf{k}}u\psi_{-\mathbf{k}'}-\int\int d\mathbf{x}d\mathbf{x}'\psi_{\mathbf{k}}(\mathbf{x})u(x)Gu(x')\psi_{-\mathbf{k}'}(\mathbf{x}')}. \tag{8}$$

Finally it is observed that, regarded as a functional of $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}'}$, the expression on the right side of Eq. (8)

occurring in the variational principle can be evaluated in closed form for reasonable trial functions such as plane waves. The possibility of this evaluation appears to stem from the relatively simple form of the Fourier transform of the Yukawa potential. For other well shapes the variational principle need not yield closed form expressions for the scattering amplitude. In particular we have been unable to evaluate the integrals in closed form for square wells (except in the forward direction). Unless the integrals can be at least approximately evaluated in closed form, computations using the Schwinger variational formulation can be so tedious as to offer insufficient or no advantage over exact numerical integration of the differential equation. L. Mower [Phys. Rev. 89, 947 (1953)] lists some special cases for which the integrals can be evaluated in closed form.

[7] E. Gerjuoy and D. S. Saxon, Phys. Rev. 85, 939 (1952).

[8] The S wave shift has been discussed in detail in reference 3 and in the more readily available article by J. M. Blatt and J. D. Jackson, Phys. Rev. 76, 18 (1949). A discussion for general L values has been given by B. P. Lippman and J. Schwinger, Phys. Rev. 79, 469 (1950) and F. Rohrlich and J. Eisenstein, Phys. Rev. 75, 705 (1949).

[9] See, e.g., H. Levine and J. Schwinger, Phys. Rev. 74, 958 (1948).

is stationary[10] with respect to arbitrary independent variations of ψ_k and $\psi_{-k'}$ about their true values as determined by Eq. (3), and the corresponding equation for $\psi_{-k'}$. Furthermore the right side of Eq. (8) is independent of the amplitudes of ψ_k and $\psi_{-k'}$.

Stationary expressions for the phase shifts are obtained in a similar fashion. The standard expansion in spherical harmonics leads to a set of independent differential equations for the partial waves $\psi_L(x)$, from which are constructed a set of integral equations incorporating the boundary conditions, namely,

$$\psi_L(x)=j_L(kx)+\int_0^\infty dx'x'^2g_L(x,x')u(x')\psi_L(x'), \quad (9)$$

where

$$g_L(x,x')=ikj_L(kx_<)h_L^{(1)}(kx_>). \quad (10)$$

Here $x_<$ denotes the smaller, $x_>$ the larger, of x and x'; and j_L and $h_L^{(1)}$ are the standard spherical Bessel and Hankel functions. It then follows that the phase shifts can be expressed in the form

$$k(\cot\delta_L-i)=\frac{1}{\left[\int_0^\infty dxx^2ju\psi_L\right]^2}\left\{\int_0^\infty dxx^2u\psi_L{}^2-\int_0^\infty\int_0^\infty dxdx'x^2x'^2\psi_L(x)u(x)g_Lu(x')\psi_L(x')\right\}\equiv Q_L. \quad (11)$$

The expression on the right side of Eq. (11) is stationary with respect to arbitrary variation of ψ_L about its true value, as determined from Eq. (9), and is independent of the amplitude of ψ_L.

We now compare some of the features of the formally equivalent results expressed by Eqs. (8) and (11). We observe first, that Eq. (11) can be derived directly from Eq. (8). This is effected by expanding the unknown functions ψ_k and $\psi_{-k'}$ in spherical harmonics. We write

$$\psi_k=\sum_{L,m}\alpha_L{}^m\psi_L(x)Y_L{}^m(\theta,\phi), \quad (12)$$

$$\psi_{-k'}=\sum_{L,m}\beta_L{}^m\psi_L(x)Y_L{}^{m*}(\theta,\phi). \quad (13)$$

Substituting in Eqs. (5)–(7) we find

$$B(\mathbf{k}\to\mathbf{k}')=\sum_{L,m}\alpha_L{}^m(-i)^LY_L{}^m(\theta_{k'},\phi_{k'})\times\int_0^\infty dxx^2j_Lu\psi_L\equiv\sum_{L,m}\alpha_L{}^mA_{1L}{}^m,$$

$$B(-\mathbf{k}'\to-\mathbf{k})=\sum_{L,m}\beta_L{}^mi^LY_L{}^{m*}(\theta_k,\phi_k)\times\int_0^\infty dxx^2j_Lu\psi_L\equiv\sum_{L,m}\beta_L{}^mA_{2L}{}^m, \quad (14)$$

$$C(\mathbf{k}\to\mathbf{k}')=\sum_{L,m}\alpha_L{}^m\beta_L{}^m\left[\frac{1}{4\pi}\int_0^\infty dxx^2u\psi_L{}^2-\frac{1}{4\pi}\int_0^\infty\int_0^\infty dxdx'x^2x'^2\times\psi_L(x)u_L(x)g_Lu_L(x')\psi_L(x')\right]\equiv\sum_{L,m}\alpha_L{}^m\beta_L{}^mA_{3L}{}^m,$$

where the quantities $A_{1L}{}^m$, $A_{2L}{}^m$, $A_{3L}{}^m$ are defined by Eq. (14).

$A(\mathbf{k}\to\mathbf{k}')$, Eq. (8), is stationary with respect to arbitrary independent variations of ψ_k and $\psi_{-k'}$, which means, using Eqs. (12) and (13) in Eq. (8), that $A(\mathbf{k}\to\mathbf{k}')$ must be stationary with respect to arbitrary independent variations of the parameters $\alpha_L{}^m$ and $\beta_L{}^m$ about their correct values. We note that

$$\delta A(\mathbf{k}\to\mathbf{k}')=\frac{B_1}{C}\delta B_2+\frac{B_2}{C}\delta B_1-\frac{B_1B_2}{C^2}\delta C, \quad (15)$$

where we have condensed the notation, with $B_1=B(\mathbf{k}\to\mathbf{k}')$, $B_2=B(-\mathbf{k}'\to-\mathbf{k})$. Using Eq. (14) in Eq. (15), and equating to zero the coefficients of $\delta\alpha_L{}^m$ and $\delta\beta_L{}^m$, we then obtain

$$\alpha_L{}^m=\frac{C}{B_2}\frac{A_{2L}{}^m}{A_{3L}{}^m}, \quad (16)$$

$$\beta_L{}^m=\frac{C}{B_1}\frac{A_{1L}{}^m}{A_{3L}{}^m}. \quad (17)$$

These variationally determined values of $\alpha_L{}^m$ and $\beta_L{}^m$, when substituted into Eq. (8), yield

$$A(\mathbf{k}\to\mathbf{k}')=\sum_L(2L+1)P_L(\cos\theta)(1/Q_L), \quad (18)$$

where θ is the angle between $\mathbf{k}'$ and $\mathbf{k}$. Moreover we observe that Eq. (18) has been deduced without making use of any of the properties of $\psi_L(x)$. Hence, by virtue of Eq. (11), Eq. (18) is seen to be equivalent to the standard form[11]

$$A(\mathbf{k}\to\mathbf{k}')=\frac{1}{2ik}\sum_L(2L+1)(e^{2i\delta_L}-1)P_L(\cos\theta), \quad (19)$$

where in Eq. (19) the phase shifts are now defined in terms of the unknown functions ψ_L, by the stationary expression Eq. (11), and are of course identical with the exact phase shifts when ψ_L are the correct radial functions. Thus in spite of their formal equivalence, there is considerable difference in the use of Eqs. (18)

[10] Some discussion of the general formal requirements for constructing stationary expressions like Eq. (8) is given in E. Gerjuoy and D. S. Saxon, "Variational principles for the acoustic field" (to be published).

[11] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), p. 105.

and (11) as the basis for actual approximations to the cross section. Obviously, if the integrals can be evaluated in closed form for suitable trial functions, Eq. (8) is much more convenient than Eq. (11), which necessarily involves tedious sums over spherical harmonics. On the other hand, Eq. (11) is inherently more accurate than Eq. (8) in the sense of the variational principle, since Eq. (11) corresponds to the use of the "best" or variationally determined set of parameters α_L^m, β_L^m in Eq. (8), for the particular set of radial trial functions $\psi_L(x)$ which is being used. Or, to put it another way, use of, say, the incident plane wave as a trial function in Eq. (11), i.e., using $\psi_L(x)=j_L(kx)$, corresponds to using in Eq. (8) not the incident wave, but rather an expression in which the plane wave expansion coefficients have been replaced by a corresponding set of variationally determined coefficients.

At this point we interpolate a remark concerning the reliability of variational methods in general as applied to scattering problems. Inasmuch as no minimum principle has been established for these problems, one cannot guarantee that the use of a sequence of trial functions containing an increasing number of variationally determined parameters will uniformly improve the estimated scattering amplitude or cross section.[12] What is desired is a trial function which approximates as closely as possible the correct wave function. When physical reasoning points to a certain trial function, this choice of trial function probably should be favored over a variationally determined trial function. In the absence of physical indications, however, it is reasonable to expect that a variationally determined trial function will not be a very bad choice, and that it is likely to be better to choose the parameters variationally than to choose them purely arbitrarily. It is in this sense that the variationally determined set of parameters α_L^m, β_L^m are said to be the "best" set.

III. PROJECTION OPERATOR FORMULATION

We now seek a somewhat different expression of the variational principle which represents a compromise between the greater complexity of Eq. (11) and the lesser accuracy of Eq. (8). We proceed by analogy to the angular momentum eigenfunction expansion which leads from Eq. (8) to Eq. (11). Namely we expand ψ_x in the eigenfunctions of operators which commute with the Hamiltonian but which are so chosen that they generate only a finite complete set. In particular, we shall choose the parity operator and some suitable finite rotation operator for this purpose. We first develop the properties of the associated projection operators, in terms of which the desired expansion is conveniently obtained.

Consider initially the parity operator P, eigenvalues ± 1. We denote the corresponding eigenfunctions by ψ_+ and ψ_- and observe that

$$P_+=\tfrac{1}{2}(1+P), \tag{20}$$

$$P_-=\tfrac{1}{2}(1-P), \tag{21}$$

are the corresponding projection operators since for any function ψ:

$$P_+\psi=\psi_+, \quad P_-\psi=\psi_-. \tag{22}$$

That is to say,

$$PP_+\psi=P_+\psi, \quad PP_-\psi=-P_-\psi. \tag{23}$$

Next, introduce the rotation operator R denoting rotation through the angle $2\pi/N$ about some arbitrary fixed axis. Then $R^N=1$ and the eigenvalues λ_n of R are

$$\lambda_n=e^{2\pi ni/N}, \quad n=0, 1, \cdots, N-1. \tag{24}$$

The corresponding projection operators are

$$P_n=\frac{1}{N}\sum_{p=0}^{N-1}\lambda_n^{N-p}R^p, \tag{25}$$

since

$$RP_n=\lambda_nP_n, \tag{26}$$

as is easily demonstrated. Also

$$\sum_{n=0}^{N-1}P_n=1. \tag{27}$$

For any function ψ we define ψ_n by

$$\psi_n=P_n\psi. \tag{28}$$

Then ψ_n is an eigenfunction of R with the eigenvalue λ_n.

In the preceding equations $R\psi(\mathbf{x})$ is understood to define a new function $\psi'(\mathbf{x})$ such that

$$\psi'(\mathbf{x})=R\psi(\mathbf{x})=\psi(R\mathbf{x}), \tag{29}$$

where $R\mathbf{x}$ is the new vector (in some specified fixed coordinate system) into which $\mathbf{x}$ rotates under R. For any integral over all solid angles (along $\mathbf{x}$), with ϕ and ψ arbitrary functions,

$$\int\phi(\mathbf{x})[R\psi(\mathbf{x})]=\int\phi(\mathbf{x})\psi(R\mathbf{x})=\int\phi(R^{-1}\mathbf{x})\psi(\mathbf{x})$$

$$=\int[R^{-1}\phi(\mathbf{x})]\psi(\mathbf{x}). \tag{30}$$

Of course the second equality in Eq. (30) is a consequence of the introduction of new rotated polar and azimuth axes, or equivalently of the introduction of new primed variables of integration defined by $\mathbf{x}'=R\mathbf{x}$, $\mathbf{x}=R^{-1}\mathbf{x}'$. Thus Eq. (30) demonstrates that the definition of $R\psi$, Eq. (29), retains the orthogonality of R,

[12] Indeed one cannot even guarantee that for a given trial function the variational principle will increase the accuracy over less sophisticated approximation techniques. Thus, for example, calculation of the Coulomb scattering cross section in Born approximation gives the correct cross section, as is well known, while use of the incident plane wave as a trial function in the variational principle does not. For further details see the discussion of this point in Sec. IV.

i.e., that

$$\bar{R}=R^{-1} \tag{31}$$

in integrals like those of Eq. (30).

From Eq. (31), using relations like $\lambda_n^{-1}=\lambda_{N-n}$, $R^{-1}=R^{N-1}$, it is simple to show

$$\bar{P}_n=P_{N-n}. \tag{32}$$

It is also true that if ϕ_n is an eigenfunction of R with eigenvalue λ_n,

$$P_{n'}\phi_n=\frac{1}{N}\sum_{p=0}^{N-1}\lambda_{n'}{}^{N-p}\lambda_n{}^p\phi_n=\frac{1}{N}\sum_{p=0}^{N-1}e^{2\pi i(n-n')/N}\phi_n$$
$$=\delta_{nn'}\phi_n, \tag{33}$$

and therefore, by Eq. (28), that

$$P_{n'}P_n=\delta_{nn'}P_n. \tag{34}$$

Hence if ϕ_n is any eigenfunction with eigenvalue λ_n, and ψ_{N-n} is defined as in Eq. (28), then, by Eqs. (32) and (33),

$$\int\psi_{N-n}\phi_n=\int[P_{N-n}\psi]\phi_n$$
$$=\int\psi[\bar{P}_{N-n}\phi_n]=\int\psi\phi_n. \tag{35}$$

Similarly, using Eq. (34), one obtains

$$\int\phi_m\phi_n=0, \quad m\neq N-n \tag{36}$$

for any two eigenfunctions ϕ_m, ϕ_n with eigenvalues λ_m, λ_n. It should be noted that in Eqs. (32)–(36) we have extended the definitions so that $\lambda_N=\lambda_0=1$, whence also $P_N=P_0$, $\psi_N=\psi_0$, etc.

We are now in a position to derive readily the desired variational expressions in terms of the simultaneous eigenfunctions of P and R. Such functions, we denote by $\phi_{n,\pm}$. They satisfy the relations

$$P\phi_{n,\pm}=\pm\phi_{n,\pm}, \quad R\phi_{n,\pm}=\lambda_n\phi_{n,\pm}, \tag{37}$$

and can be constructed from any arbitrary function ψ via the defining equation

$$P_\pm P_n\psi=P_nP_\pm\psi=\psi_{n,\pm}. \tag{38}$$

In analogy to Eqs. (12) and (13), we write

$$\psi_{\mathbf{k}}=\sum_n\sum_{i=\pm}\alpha_{n,i}\phi_{n,i}{}^{\mathbf{k}},$$
$$\psi_{-\mathbf{k}'}=\sum_n\sum_{i=\pm}\beta_{n,i}\phi_{n,i}{}^{-\mathbf{k}'}, \tag{39}$$

where the superscripts indicate that $\phi_{n,i}{}^{\mathbf{k}}$ and $\phi_{n,i}{}^{-\mathbf{k}'}$ need not be the same functions. Corresponding to Eq. (14), we have:

$$B(\mathbf{k}\to\mathbf{k}')=\sum_{n,i}\frac{1}{4\pi}\alpha_{n,i}\int d\mathbf{x}e^{-i\mathbf{k}'\cdot\mathbf{x}}u(x)\phi_{n,i}{}^{\mathbf{k}}$$
$$\equiv\sum_{n,i}\alpha_{n,i}A_{n,i}{}^{(1)},$$
$$B(-\mathbf{k}'\to-\mathbf{k})=\sum_{n,i}\frac{1}{4\pi}\beta_{n,i}\int d\mathbf{x}e^{i\mathbf{k}\cdot\mathbf{x}}u(x)\phi_{n,i}{}^{-\mathbf{k}'}$$
$$\equiv\sum_{n,i}\beta_{n,i}A_{n,i}{}^{(2)}, \tag{40}$$
$$C(\mathbf{k}\to\mathbf{k}')=\sum_{n,i}\alpha_{n,i}\beta_{N-n,i}\Big[\frac{1}{4\pi}\int d\mathbf{x}\phi_{n,i}{}^{\mathbf{k}}u(x)\phi_{N-n,i}{}^{-\mathbf{k}'}$$
$$-\frac{1}{4\pi}\int d\mathbf{x}d\mathbf{x}'\phi_{n,i}{}^{\mathbf{k}}(\mathbf{x})u(x)$$
$$\times G(\mathbf{x},\mathbf{x}')u(x')\phi_{N-n,i}{}^{-\mathbf{k}'}(\mathbf{x}')\Big]$$
$$\equiv\sum_{n,i}\alpha_{n,i}\beta_{N-n,i}A_{n,i}{}^{(3)}.$$

The expression for $C(\mathbf{k}\to\mathbf{k}')$ in Eq. (40) is justified with the aid of Eqs. (32), (33), and (36), using obvious properties of the parity operator and noting that

$$R(\mathbf{x})R(\mathbf{x}')G(\mathbf{x},\mathbf{x}')=G(\mathbf{x},\mathbf{x}'), \tag{41}$$

where $R(\mathbf{x})$ rotates $\mathbf{x}$ only, all of which yields

$$P_{N-n}(\mathbf{x})G=N^{-1}\textstyle\sum_p\lambda_{N-n}{}^{N-p}R^p(\mathbf{x})G$$
$$=N^{-1}\textstyle\sum_p\lambda_{N-n}{}^{N-p}R^{-p}(\mathbf{x}')G=P_n(\mathbf{x}')G. \tag{42}$$

We require $A(\mathbf{k}\to\mathbf{k}')$, Eq. (8), to be stationary with respect to arbitrary independent variations of the parameters $\alpha_{n,i}$ and $\beta_{n,i}$. Then, using Eq. (15) we find that, for variationally determined values of $\alpha_{n,i}$ and $\beta_{n,i}$,

$$A(\mathbf{k}\to\mathbf{k}')=\sum_n\sum_{i=\pm}A_{n,i}(\mathbf{k}\to\mathbf{k}'),$$
$$A_{n,i}(\mathbf{k}\to\mathbf{k}')=A_{n,i}{}^{(1)}A_{N-n,i}{}^{(2)}/A_{n,i}{}^{(3)}. \tag{43}$$

This derivation of Eq. (43) is wholly analogous to the derivation of Eq. (11) from Eq. (8), for arbitrary ψ_L in Eq. (12), via variation of $\alpha_L{}^m$ and $\beta_L{}^m$. Alternatively we can use the expansions Eq. (39) in Eq. (1) to obtain

$$\nabla^2\phi_{n,\pm}{}^{\mathbf{k}}+[k^2+u(x)]\phi_{n,\pm}{}^{\mathbf{k}}=0, \tag{44}$$

and similarly for $\phi_{n,\pm}{}^{-\mathbf{k}'}$. The function $\phi_{n,\pm}{}^{\mathbf{k}}$ obeys the integral equation

$$\phi_{n,\pm}{}^{\mathbf{k}}(\mathbf{x})=P_nP_\pm e^{i\mathbf{k}\cdot\mathbf{x}}+\int d\mathbf{x}'G(\mathbf{x},\mathbf{x}')u(x')\phi_{n,\pm}{}^{\mathbf{k}}(\mathbf{x}'). \tag{45}$$

From Eq. (45) and the corresponding equation for $\phi_{n,\pm}{}^{-\mathbf{k}'}$, it is not difficult to determine that Eq. (43) is valid for the exact scattering amplitude, for which in fact

$$A_{n,i}(\mathbf{k}\to\mathbf{k}')=A_{n,i}{}^{(1)}(\mathbf{k}\to\mathbf{k}')$$
$$=A_{N-n,i}{}^{(2)}(\mathbf{k}\to\mathbf{k}')=A_{n,i}{}^{(3)}(\mathbf{k}\to\mathbf{k}'), \tag{46}$$

when, in Eq. (40), $\phi_{n,i}^{\mathbf{k}}$ and $\phi_{n,i}^{-\mathbf{k}'}$ take on their true values, i.e., $\phi_{n,i}^{\mathbf{k}}$ satisfies Eq. (45). Moreover, the expression for $A_{n,i}(\mathbf{k}\to\mathbf{k}')$, Eq. (43) is stationary with respect to arbitrary variation of $\phi_{n,i}^{\mathbf{k}}$ and $\phi_{n,i}^{-\mathbf{k}'}$ about their true values. This second derivation of Eq. (43) is of course analogous to the customary derivation of Eq. (11) from Eq. (9).

Equation (43) is the variational formulation we have been seeking. It can be expressed in a simpler form by the following device. We first specify that the functions $\phi_{n\pm}^{\mathbf{k}}$ and $\phi_{n,\pm}^{-\mathbf{k}'}$ are constructed from given trial functions $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}'}$ by means of Eq. (38). We make the further restriction that $\psi_{\mathbf{k}}(\mathbf{x})$ is a function only of the angle between $\mathbf{k}$ and $\mathbf{x}$, i.e., that it does not depend on the azimuth of $\mathbf{x}$ about $\mathbf{k}$. This condition is certainly satisfied for the true $\psi_{\mathbf{k}}(\mathbf{x})$ which solves Eq. (3). Similarly, $\psi_{-\mathbf{k}'}(\mathbf{x})$ is independent of the azimuth of $\mathbf{x}$ relative to $-\mathbf{k}'$. This means that for any rotation R, in the notation of Eq. (41),

$$R(\mathbf{k})R(\mathbf{x})\psi_{\mathbf{k}}(\mathbf{x})=\psi_{\mathbf{k}}(\mathbf{x}), \tag{47}$$

and therefore just as Eq. (42) followed from Eq. (41)

$$P_n(\mathbf{x})\psi_{\mathbf{k}}(\mathbf{x})=P_{N-n}(\mathbf{k})\psi_{\mathbf{k}}(\mathbf{x}). \tag{48}$$

From Eq. (48), and also Eq. (42) we infer that

$$\begin{aligned}A_{n,\pm}{}^{(1)}&=\frac{1}{4\pi}\int d\mathbf{x}e^{-i\mathbf{k}'\cdot\mathbf{x}}u(x)[P_n(\mathbf{x})P_{\pm}(\mathbf{x})\psi_{\mathbf{k}}(\mathbf{x})]\\&=P_{N-n}(\mathbf{k})P_{\pm}(\mathbf{k})B(\mathbf{k}\to\mathbf{k}'),\\A_{N-n,\pm}{}^{(2)}&=\frac{1}{4\pi}\int d\mathbf{x}e^{i\mathbf{k}\cdot\mathbf{x}}u(x)[P_{N-n}(\mathbf{x})P_{\pm}(\mathbf{x})\psi_{-\mathbf{k}'}(\mathbf{x})]\\&=\frac{1}{4\pi}\int d\mathbf{x}[P_n(\mathbf{x})P_{\pm}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}}]u(x)\psi_{-\mathbf{k}'}(\mathbf{x})\\&=P_{N-n}(\mathbf{k})P_{\pm}(\mathbf{k})B(-\mathbf{k}'\to-\mathbf{k}),\end{aligned} \tag{49}$$

and, by slightly more complicated manipulations, using Eq. (34) as well, that

$$A_{n,\pm}{}^{(3)}=P_{N-n}(\mathbf{k})P_{\pm}(\mathbf{k})C(\mathbf{k}\to\mathbf{k}'), \tag{50}$$

where $B(\mathbf{k}\to\mathbf{k}')$, $B(-\mathbf{k}'\to-\mathbf{k})$, and $C(\mathbf{k}\to\mathbf{k}')$ are defined by Eqs. (5)–(7).

Equations (49) and (50) make Eq. (43) particularly convenient, since they merely involve linear combinations of the same functions B and C which enter into the more familiar stationary expression Eq. (8). We further note that the extension of our result to the situation in which space exchange forces are involved is trivial, since space exchange merely introduces different interactions in even and odd states. The point here is of course that Eq. (43) is already decomposed into sums over even and odd states. Hence one need only use the corresponding even and odd interactions in computing the contribution of each to the total scattering amplitude.[13] Thus, for example, for the Serber type of exchange, wherein the interaction vanishes in odd states, Eq. (43) becomes simply

$$A(\mathbf{k}\to\mathbf{k}')=\textstyle\sum_n A_{n,+}(\mathbf{k}\to\mathbf{k}'). \tag{51}$$

To this point, R has representated a rotation about a fixed but arbitrary axis. In the actual application of Eq. (43), together with Eqs. (49) and (50), it is necessary to make some definite choice of the orientation of this axis, with respect to $\mathbf{k}$ for instance. For the true solutions $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}'}$ the results are entirely independent of the orientation of this axis of rotation. In this event, $B(\mathbf{k}\to\mathbf{k}')=B(-\mathbf{k}\to-\mathbf{k})=C(\mathbf{k}\to\mathbf{k}')$ so that, using Eqs. (49) and (50) in Eq. (43), and noting that $P_{\pm}$ obviously sums to unity, as does P_{N-n}, by Eq. (27), we have, for the true $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}'}$,

$$A(\mathbf{k}\to\mathbf{k}')=\sum_{n,i}P_{N-n}(\mathbf{k})P_i(\mathbf{k})B(\mathbf{k}\to\mathbf{k}')=B(\mathbf{k}\to\mathbf{k}'), \tag{52}$$

and $B(\mathbf{k}\to\mathbf{k}')$ does not involve R.

When $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}'}$ are merely some suitable trial functions however, the results using Eq. (43) do depend on the axis of rotation of R. As an illustration, if the axis of rotation is chosen to lie along $\mathbf{k}$, the quantities B and C in Eqs. (49) and (50) are seen to be eigenfunctions of $R(\mathbf{k})$ with eigenvalue unity. Then only the terms with $n=0$ fail to vanish in Eq. (43), recalling Eq. (33), and we are left with the not unexpected result that with this choice of axis

$$A(k\to k')=A_+(k\to k')+A_-(k\to k'), \tag{53}$$

where A_+ and A_-, the even and odd parts of the scattering amplitude, are given by

$$A_{\pm}=\frac{[P_{\pm}(\mathbf{k})B(\mathbf{k}\to\mathbf{k}')][P_{\pm}(\mathbf{k})B(-\mathbf{k}'\to-\mathbf{k})]}{P_{\pm}(\mathbf{k})C(\mathbf{k}\to\mathbf{k}')}. \tag{54}$$

On the other hand, suppose R corresponds to a rotation through an angle of $\frac{2}{3}\pi$ about an axis making equal direction cosines with the three rectangular axes. In other words, when operating on the coordinate axes, R causes a cyclic interchange of these axes. Thus, if (x,y,z) are the components of a vector,

$$\begin{aligned}&Rf(x,y,z)=f(z,x,y);\\&\text{and}\\&R^3=1,\quad \lambda_n=e^{i\pi n},\quad n=0,\,1,\,2.\end{aligned} \tag{55}$$

Further, suppose we choose the z axis along $\mathbf{k}$ and take $\mathbf{k}'$ to lie in the positive x half of the x-z plane, which in a rather complicated fashion specifies the orientation relative to $\mathbf{k}$ and $\mathbf{k}'$ of the axis of rotation of R. With $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}'}$ as previously restricted, i.e., with $\psi_{\mathbf{k}}(\mathbf{x})$ independent of the azimuth of $\mathbf{x}$ about $\mathbf{k}$, it is intuitively obvious, and can be proved formally directly from Eqs. (5)–(7) using Eqs. (31), (41), and (47), that $B(\mathbf{k}\to\mathbf{k}')$, $B(-\mathbf{k}'\to-\mathbf{k})$, and $C(\mathbf{k}\to\mathbf{k}')$ are independent of the azimuth of $\mathbf{k}'$ about $\mathbf{k}$. Hence, insofar as rotations are concerned, these quantities can be written as $f(\mathbf{k},\mathbf{k}')=f(\theta)$, with θ the scattering angle. Then for the

[13] One could arrive at the same result by inserting different even and odd interactions $u(x)$ in Eq. (44).

present choice of coordinate axes and of R, $R(\mathbf{k})f(\theta)=f(|\frac{1}{2}\pi-\theta|)$, $R^2(\mathbf{k})f(\theta)=f(\frac{1}{2}\pi)$, $P(\mathbf{k})f(\theta)=f(\pi-\theta)$, and we have

$$P_n=\tfrac{1}{3}\sum_{p=0}^{2}e^{-2\pi inp/3}R^p,$$

$$P_n(\mathbf{k})P_\pm(\mathbf{k})f(\theta)=\tfrac{1}{3}[f_\pm(\theta)+e^{-2\pi in/3}f_\pm(|\tfrac{1}{2}\pi-\theta|)+e^{2\pi in/3}f_\pm(\tfrac{1}{2}\pi)],\quad n=0,1,2 \quad (56)$$

$$f_\pm(\theta)=\tfrac{1}{2}[f(\theta)\pm f(\pi-\theta)].$$

Evidently, Eq. (56) combined with Eqs. (43), (49), and (50) generally does not yield the same result as Eq. (53).

IV. APPLICATION TO THE YUKAWA POTENTIAL

Procedure

We now describe the specific calculations we have performed. We have used the interaction

$$u(x)=u_0e^{-x}/x, \quad (57)$$

with $u_0=2.365$, which corresponds to the Yukawa well depth and range quoted in Sec. I. Scattering amplitudes have been computed at values of $k=0.6630$, 1.048, 1.406, 1.624, 1.816, corresponding to laboratory energies of 20, 50, 90, 120, 150 Mev. In the various variational formulations, the trial functions $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}'}$ are always

$$\psi_{\mathbf{k}}(\mathbf{x})=\exp(iK\mathbf{n}\cdot\mathbf{x}),\quad \psi_{-\mathbf{k}'}(\mathbf{x})=\exp(-iK\mathbf{n}'\cdot\mathbf{x}), \quad (58)$$

where $\mathbf{n}$ is the unit vector along the incident wave vector $\mathbf{k}$, and $\mathbf{n}'$ is the unit vector along the direction of the scattered wave $\mathbf{k}'$. Thus we use plane-wave trial functions, but do not necessarily equate the effective or average wave number K to the wave number k at infinity. This special choice of trial function is physically reasonable and permits evaluation in closed form of the integrals which occur. In fact, with Eqs. (57) and (58), Eqs. (5) and (6) readily yield

$$B(\mathbf{k}\to\mathbf{k}')=B(-\mathbf{k}'\to-\mathbf{k})=u_0g(K,k,\theta), \quad (59)$$

where

$$g(K,k,\theta)=1/(1+k^2+K^2-2kK\cos\theta), \quad (60)$$

with θ, the scattering angle, the angle between $\mathbf{n}$ and $\mathbf{n}'$. The only complicated integral is the one containing the Green's function in Eq. (7), which can be evaluated for arbitrary K by the same method as used by Jost and Pais[9] and by Dalitz[14] for the special case $K=k$, as well as by standard techniques, without introduction of auxiliary variables.[15] It is found that

$$C(\mathbf{k}\to\mathbf{k}')=u_0g(K,K,\theta)-u_0^2G(K,k,\theta), \quad (61)$$

where g is as in Eq. (60), and where

$$G(K,k,\theta)=\frac{1}{2K\sqrt{S}\sin(\theta/2)}\times\left\{\tan^{-1}\left(\frac{K(1+K^2-k^2)\sin(\theta/2)}{\sqrt{S}}\right)+\frac{i}{2}\log\frac{\sqrt{S}+2kK\sin(\theta/2)}{\sqrt{S}-2kK\sin(\theta/2)}\right\},$$
$$S\equiv(1+K^2+k^2)^2-4k^2K^2\cos^2(\theta/2). \quad (62)$$

In Eq. (62), $\sqrt{S}>0$ and the arctangent lies between $-\pi/2$ and $\pi/2$.

Perhaps the most consistent variational treatment would be to determine K variationally[16] at each scattering angle. However, this would require much additional numerical work, considerably more so than appears to be justifiable, see our remark at the end of Sec. II. Hence we have confined our calculations to the following three choices of the effective wave number K:

$$\text{(a)}\ K^2=k^2\equiv K_a^2,$$
$$\text{(b)}\ K^2=k^2+\tfrac{1}{3}u_0\equiv K_b^2, \quad (63)$$
$$\text{(c)}\ K^2=k^2+\tfrac{1}{2}u_0\equiv K_c^2.$$

The first is just the free-particle wave number, of course. The second is determined variationally from the requirement that the forward scattering cross section in the high-energy limit be stationary with respect to K. This is, at any rate, a high-energy variational determination of K, and is obtained without excessive manipulation. To be specific, we insert Eqs. (59)–(62) in Eq. (8) with $\theta=0$, let $K^2=k^2+\beta$, β finite, and equate to zero the derivative with respect to K in the limit that $k\to\infty$. The resultant Eq. (63b) is the same as would be found from the more complicated formulation Eq. (43), since in the high-energy limit the quantities B and C have sharp peaks in the forward direction, and are small at other angles. Consequently in $P_{N-n}(\mathbf{k})P_\pm(\mathbf{k})B(\mathbf{k}\to\mathbf{k}')$ for example, all terms in R and P are negligible compared to the term in R^0P^0, i.e., in this limit, at this angle, $P_{N-n}(\mathbf{k})P_\pm(\mathbf{k})$ is replaceable by $1/2N$, and Eq. (43) becomes identical with Eq. (8). Equation (63c) is determined from the condition

$$\int dx u(x)e^{-i\mathbf{K}\cdot\mathbf{x}}[\nabla^2+k^2+u(x)]e^{i\mathbf{K}\cdot\mathbf{x}}=0. \quad (64)$$

Equation (64) is an expression of the requirement that the trial function $e^{i\mathbf{K}\cdot\mathbf{x}}$, which does not satisfy the

[14] R. H. Dalitz, Proc. Roy. Soc. (London) **A206**, 509 (1951).

[15] After transformation to momentum space, with $\mathbf{n}'$ as polar axis and the azimuth axis in the plane of $\mathbf{n}$, $\mathbf{n}'$, integration over azimuth and thence over the polar angle can be effected by successive use of formulas 300 and 195, of B. O. Peirce, *A Short Table of Integrals* (Ginn and Company, Boston, 1910). The resultant integral, though complicated in form, can be evaluated by closing the contour in the complex plane, taking care to avoid circling the branch points of the integrand.

[16] The scattering amplitude is complex and with a single variational parameter, as with K, in the present instance, it generally is not possible to cause the derivatives of the real and imaginary parts to vanish simultaneously. Thus it would be necessary to decide say that the magnitude of $A(\mathbf{k}\to\mathbf{k}')$ is to be made stationary, without regard to phase.

Schrödinger equation anywhere, at least satisfy the Schrödinger equation on an average weighted by the Yukawa potential, this weighted average favoring the region of space, where the integrands in B and C are significant.[17]

Three successively more complicated variational formulations have been employed, namely:

I. Equation (8), of course with the trial functions of Eq. (58).

II. Equation (43) in the form of Eqs. (53) and (54). This involves a decomposition into even and odd terms, and corresponds to using in Eq. (8) a trial function which is a linear combination of forward and backward traveling waves, with variationally determined amplitudes.

III. Equation (43) combined with the results of Eqs. (49), (50), and (56). This involves a further decomposition into the eigenfunctions of the rotation operator R of Eq. (55), and corresponds to using in Eq. (8) a trial function which is a linear combination of waves traveling in both directions along three mutually perpendicular axes including the forward and backward directions, with variationally determined amplitudes.

As with K, so in formulation III the orientation angles of the axis of rotation of R should presumably be regarded as variational parameters, to be determined for fixed N and K so as to make the scattering amplitude stationary at each scattering angle. Again the extra work required does not seem worth while, and we have fixed the orientation of the rotation axis mainly for reasons of convenience. Since according to Eq. (52) the results are independent of the orientation of the rotation axis when $\psi_{\mathbf{k}}$ and $\psi_{-\mathbf{k}'}$ are the true solutions, it is likely that the results will not depend strongly on the orientation of this axis provided the trial functions Eq. (58) are "reasonable." In any event this choice of rotation axis leads to good results, which only should be improved by a better founded choice of axis. It may seem that the cross sections using variational formulation III will be azimuth dependent, since we have specifically assumed $\mathbf{k}'$ lies in the $x-z$ plane in deriving Eq. (56), and in fact formulation III does correspond to the use of an azimuth dependent trial function in Eq. (8). The dependence on azimuth is only apparent however. We always keep $\mathbf{k}'$ in the positive x half of the $x-z$ plane, or in other words, we change the axis of rotation of R, Eq. (58), as the azimuth of $\mathbf{k}'$ changes, thereby retaining independence of the azimuth of $\mathbf{k}'$ relative to $\mathbf{k}$ in our final answer.

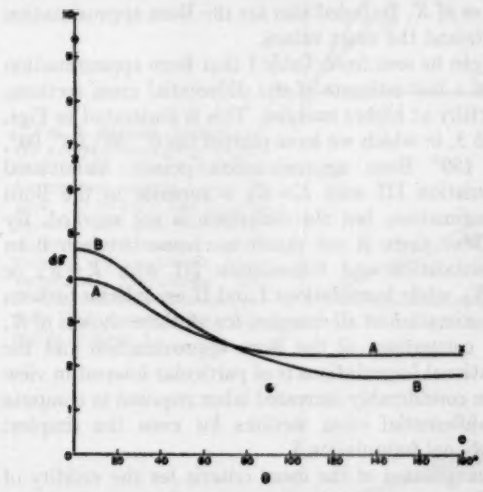

FIG. 1. Differential cross section at 20 Mev vs scattering angle. Curve A is the exact result, curve B the result according to variational formulation III, $K=K_b$. The circles and crosses are points computed respectively according to formulation I and II, $K=K_b$.

[17] Strictly speaking, Eqs. (63b) and (63c), which have been calculated on the assumption of ordinary forces, should be modified for exchange forces, but we shall not worry about this. They are merely reasonable guesses as to the best choice of K and, as will be seen, they do yield good estimates of the cross section when exchange forces are included.

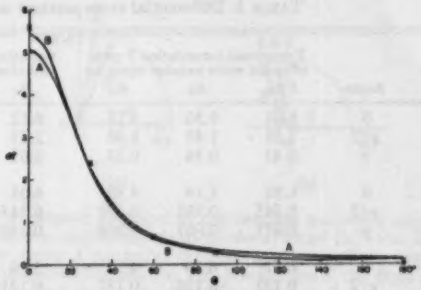

FIG. 2. Differential cross section at 90 Mev vs scattering angle. Curve A is the exact result, curve B the result according to variational formulation III, $K=K_b$. The crosses are points computed according to Born approximation.

Variational and Born Results

Our numerical results definitely show that the variational formulation III is superior to the simpler formulations I and II, and that formulation II is itself an improvement over the more customary formulation I. By and large, the choice $K=K_b$ is the best of the three we have tried, although at lower energies $K=K_a$ may be superior. In Figs. 1–3 we have plotted the differential cross sections $d\sigma$ computed according to formulation III, with $K=K_b$, at energies of 20 Mev, 90 Mev, and 150 Mev, including ordinary forces only, and, on the same graphs, display for comparison the exact differential cross sections[4] at the same energies. The differential cross sections, in these and subsequent figures, are given in the units a^2 per steradian, where a is the length introduced in Eqs. (1) and (57). In Fig. 1 we have plotted also the 0°, 90°, and 180° differential cross sections which are obtained from formulations I and II, with the same value of $K=K_b$. With the simple formulation I the variational values are much too high at 0° and much too low at 180°. Formulation II mixes

TABLE I. Differential cross sections vs energy and angle as calculated by different methods.

Energy (Mev)	Angle	Variational formulation I with effective wave number equal to K_a	K_b	K_c	Variational formulation II with effective wave number equal to K_a	K_b	K_c	Variational formulation III with effective wave number equal to K_a	K_b	K_c	Born	Exact
20	0	8.61	9.36	3.73	6.12	6.70	3.79	4.93	4.64	3.13	5.60	3.99
	$\pi/2$	1.51	1.49	1.36	1.51	1.49	1.36	1.89	2.14	2.26	1.58	2.28
	π	0.41	0.34	0.25	2.01	2.41	1.48	1.40	1.78	2.06	0.74	2.27
50	0	6.81	7.19	4.38	6.05	6.39	4.13	5.09	5.08	3.25	5.60	4.58
	$\pi/2$	0.345	0.353	0.372	0.345	0.353	0.372	0.561	0.659	0.757	0.546	0.752
	π	0.071	0.067	0.064	0.525	0.591	0.467	0.276	0.359	0.501	0.192	0.531
90	0	6.27	6.50	4.74	5.99	6.21	4.59	5.34	5.42	3.95	5.60	5.07
	$\pi/2$	0.121	0.125	0.137	0.121	0.125	0.137	0.248	0.283	0.333	0.227	0.309
	π	0.023	0.023	0.024	0.191	0.206	0.181	0.093	0.115	0.162	0.071	0.151
120	0	6.10	6.28	4.90	5.94	6.11	4.80	5.44	5.54	4.31	5.60	5.25
	$\pi/2$	0.072	0.074	0.081	0.072	0.074	0.081	0.163	0.183	0.214	0.142	0.190
	π	0.013	0.014	0.014	0.112	0.118	0.108	0.056	0.067	0.090	0.028	0.048
150	0	6.00	6.14	5.01	5.90	6.04	4.93	5.50	5.60	4.56	5.60	5.31
	$\pi/2$	0.048	0.049	0.054	0.048	0.049	0.054	0.117	0.129	0.149	0.097	0.127
	π	0.009	0.009	0.010	0.073	0.076	0.071	0.038	0.044	0.057	0.028	0.048

at 0° the 0° and 180° amplitudes of I, and similarly at 180°, with a consequent significant decrease in the variational estimate of $d\sigma$ at 0°, and a corresponding increase at 180°. The 90° point of formulation I is of course unaltered since the odd part of the scattering amplitude vanishes at 90°. By introducing the additional partial waves of formulation III we produce additional mixing at 90° and at other angles, and further improve the variational estimate of the angular distribution. In fact, in Fig. 1, the 90° value of formulation III is very nearly correct, but the variational estimate is still too high at low angles and too low at large angles. Presumably the introduction of more partial waves, with the more complicated mixing which would result, would produce further improvement in the differential cross section until finally we expand in the infinite set Y_L^m and attain the phase shift formulation of Eq. (11).

The relative merits of these three variational formulations are maintained at higher energies, but the differences between them become less marked in the forward peak, which makes the principal contribution to the total cross section. This is so because, just as explained in the discussion of Eq. (63b), at angles near 0° the large forward peak reduces the effectiveness of the mixing introduced by the projection operators. At wider angles the added partial waves of II and then III continue to improve the results. The correctness of these remarks and of the assertions in the preceding paragraph, can be gauged from Table I, which summarizes the 0°, 90°, and 180° results at the five energies for the three variational formulations and the three choices of K. Included also are the Born approximation results and the exact values.

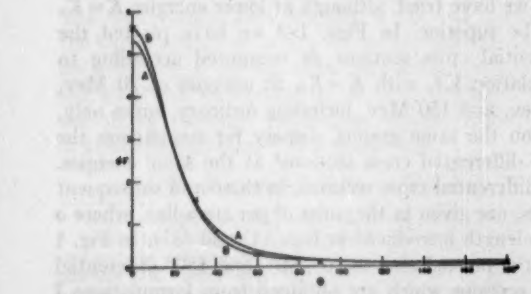

FIG. 3. Differential cross section at 150 Mev vs scattering angle. Curve A is the exact result, curve B the result according to variational formulation III, $K=K_b$. The crosses are points computed according to Born approximation.

It can be seen from Table I that Born approximation yields a fair estimate of the differential cross sections, especially at higher energies. This is illustrated in Figs. 2 and 3, in which we have plotted the 0°, 30°, 60°, 90°, and 180° Born approximation points. Variational formulation III with $K=K_b$ is superior to the Born approximation, but the difference is not marked. By 150 Mev there is not much to choose between Born approximation and formulation III with $K=K_a$ or $K=K_b$, while formulations I and II are inferior to Born approximation at all energies, for all three choices of K. This comparison of the Born approximation and the variational formulations is of particular interest in view of the considerably increased labor required to compute the differential cross sections for even the simplest variational formulation I.

Examination of the usual criteria for the validity of the Born approximation, which may be written as

$$u_0 \ll k^2, \quad u_0/k < 1, \tag{65}$$

in our notation,[18] shows that these are not satisfied even at 150 Mev where $k=1.816$, while $u_0=2.365$. Conse-

[18] The second of Eq. (65) expresses the requirement that across the scattering region the phase difference between the actual and free propagation be small. D. Bohm, *Quantum Theory* (Prentice-Hall, Inc., New York, 1951), pp. 553 ff.

quently, the good estimates of the differential cross sections yielded by the Born approximation seem rather accidental, and we anticipate that the phase of the complex scattering amplitude is not as happily estimated as is its magnitude.[19] This is demonstrated in Figs. 4 and 5 in which the computed scattering amplitudes are drawn in the complex plane. In each Argand diagram the circle is drawn with a radius equal to the exact scattering amplitude. The Born amplitude lies always on the real axis, as it must for any potential, while the variational and exact amplitudes are complex. The variational amplitudes in both figures are computed using formulation III, with $K=K_b$. In Fig. 4 we show the forward scattering amplitudes at 20, 90, and 150 Mev, and in Fig. 5 we plot the 90-Mev amplitudes at the angles $\pi/2$ and π. We observe that the phase of the exact scattering amplitude is in every case rather different from zero, and that this phase is much better estimated by the variational method than by the Born, although the magnitudes of the amplitudes are much the same, as we have already seen. We note also that the exact phase becomes smaller as the energy increases, and that the exact phase is larger at wide angles than at 0°. Similar figures at other energies share the same characteristics.

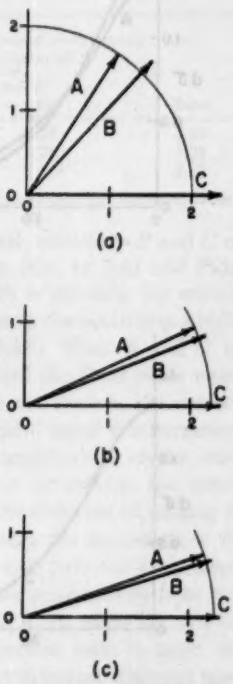

FIG. 4. Argand diagrams of the complex scattering amplitude in the forward direction for various energies: (a) 20 Mev, (b) 90 Mev, and (c) 150 Mev. In each case the vector A represents the exact amplitude, B the amplitude from variational formulation III, $K=K_b$, and C the Born amplitude. The arcs are drawn with radius equal to A.

[19] It is reasonable to attribute the good estimates of the Born approximation to the similarity of the Yukawa to the Coulomb potential. For the latter, the Born approximation happens to give the correct cross section but does not give the essentially infinite phase. The validity of Born approximation for a Yukawa well has been discussed by Jost and Pais and by Dalitz in the references cited. Dalitz also considers the transition to the Coulomb limit.

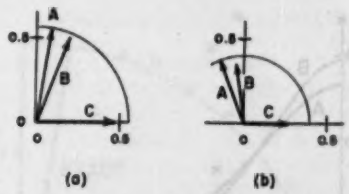

FIG. 5. Argand diagrams of the complex scattering amplitude at 90 Mev at scattering angles (a) $\theta=\pi/2$ and (b) $\theta=\pi$. In each case the vector A represents the exact amplitude, B the amplitude from variational formulation III, $K=K_b$, and C the Born amplitude. The arcs are drawn with radius equal to A.

At a scattering angle θ space exchange forces may be regarded as producing a mixture of the nonexchange amplitudes at θ and $\pi-\theta$. Since the Born amplitudes are always in phase, whereas the exact amplitudes have different phases at different scattering angles, it is apparent that it generally is not possible to have equally good results with the Born approximation for all combinations of ordinary and exchange forces. In the Yukawa case we have been discussing therefore, where Born approximation gives such good results for ordinary forces, we expect rather poorer results when exchange forces are introduced. In Figs. 6 and 7 we have plotted the exact and variational results at 90 and 150 Mev for two special combinations of ordinary and exchange forces, namely Serber forces in which the well is multiplied by P_+, and "anti-Serber" forces in which the well is multiplied by P_-. In the former case only the even waves contribute to the scattering, Eq. (51), while in the latter case we include correspondingly only the odd waves. We have used variational formulation III, with $K=K_b$, and have plotted also the Born results at 0°, 30°, 60°, and 90°. Figures 6 and 7 may be compared with the corresponding Figs. 2 and 3. Evidently, both the Serber and anti-Serber Born differential cross sections are relatively less accurate than were the Born estimates for ordinary forces, whereas the variational calculation remains about equally good for all three different types of forces. In fact for the odd scattering the variational calculation is practically indistinguishable from the exact at 90° and 150° Mev. Of course the variational calculation continues to give a good estimate of the phase as compared to the Born approximation, as is illustrated in Fig. 8 for the even forward scattering amplitude at 20 Mev. Figure 8 is constructed in the same fashion as Figs. 4 and 5, with the Born phase real, as always. We also use Fig. 8 to supplement Fig. 1 by showing how the phase as well as the amplitude is improved by adding more partial waves, and also compare the three different choices of effective wave number K. To avoid excessive complication of the figure, only formulations II and III are included in Fig. 8. We observe that both variational formulations and all choices of K give much better estimates of the phase than does Born approximation, even though the Born estimate of the magnitude of the amplitude is

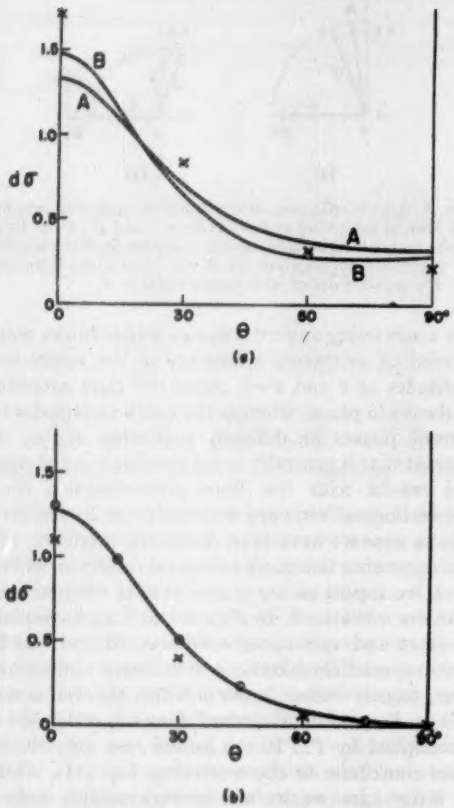

FIG. 6. (a) Even scattering cross section at 90 Mev *vs* scattering angle. Curve *A* is the exact result, curve *B* the result according to variational formulation III, $K=K_b$. The crosses are points computed according to Born approximation. (b) Odd scattering cross section at 90 Mev *vs* scattering angle. The curve represents the exact results, the circles are points computed according to variational formulation III, $K=K_b$, and the crosses are points computed according to Born approximation.

very good and certainly better than formulation II for $K=K_a$ and K_b. We see also that for the even scattering at 20 Mev, $K=K_e$ gives a better final estimate of the phase than does $K=K_b$. This is so for the even scattering at other energies as well, but the odd scattering phases definitely favor $K=K_b$, the differing behavior of the even and odd waves being especially marked at low energies. These effects can be understood. Especially at low energies the even scattering, which includes *S* wave scattering, arises from a region of space on the average closer to the origin than does the odd scattering, for which the wave function vanishes at the origin. Thus the effective wave number *K* for the even scattering is larger than for the odd scattering, especially at low energies. It is by considerations of this kind that a better choice of effective wave number can be made. For example, we have calculated the differential cross sections at 20 Mev, ordinary forces, formulation III for two modified choices of wave number, as follows. The even waves are calculated both times with $K=K_e$, but for the odd waves we calculate in one case with $K=K_a$, and in the other case with $K=K_b$. In Table II the results for these two modified choices are tabulated at 0°, 90°, and 180°, along with the exact values from Table I. The choice $K=K_b$ for the odd waves is seen to be superior, but both modifications are good. This use of different effective wave numbers for even and odd states does produce an over-all improvement, as may be seen by comparison of Table II with the 20-Mev variational formulation III section of Table I.

Utility of Variational Principle

We may infer correctly from Fig. 8, which is quite typical, that the variational method, even in its least sophisticated formulation I, with $K=K_a=k$, generally will give a better estimate of the phase of the scattering amplitude than does Born approximation. For example, at 20 Mev the correct phase angle, using ordinary forces, is 57°, while formulation I with $K=k$ is 45°. Thus in a very real sense for an arbitrary potential,

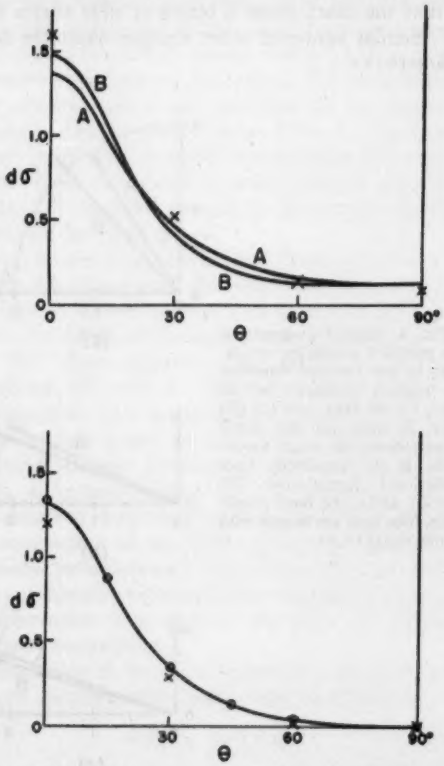

FIG. 7. (a) Even scattering cross section at 150 Mev *vs* scattering angle. Curve *A* is the exact result, curve *B* the result according to variational formulation III, $K=K_b$. The crosses are points computed according to Born approximation. (b) Odd scattering cross section at 150 Mev *vs* scattering angle. The curve represents the exact results, the circles are points computed according to variational formulation III, $K=K_b$, and the crosses are points computed according to Born approximation.

the variational method is an improvement on Born approximation, in that the variational method gives an estimate of both the real and imaginary parts of the scattering amplitude. The Born, on the other hand, estimates merely the magnitude of the amplitude, or perhaps its real part (it is not altogether clear which). It happens in the Yukawa case that the first Born approximation is very close to the correct magnitude while simultaneously the second Born correction term is large. Since for cross sections we are interested only in the magnitude of the scattering amplitude, and not in its phase, the present case evidently is especially favorable to the Born approximation.

At the same time, as we proceed to explain, it is unfavorable to the variational principle. Let us consider formulation I with $K=k$, i.e., with Born trial functions. Under these circumstances, by Eqs. (59) and (61),

$$B=u_0 g(k,k,\theta), \quad C=u_0 g(k,k,\theta)-u_0^3 G(k,k,\theta), \quad (66)$$

where $u_0 g(k,k,\theta)$ is precisely the Born approximation scattering amplitude. For the correct wave function B and C must be equal. Thus a measure of the closeness to the correct function of a given trial function is the difference between B and C. Now since $G(k,k,\theta)$ is complex whereas $g(k,k,\theta)$ is real, obviously B and C of Eq. (66) cannot be equal. In fact, as Jost and Pais[2] have shown, $u_0^3 G(k,k,\theta)$, which is precisely the second Born approximation correction to the scattering amplitude, is comparable to $u_0 g(k,k,\theta)$. Thus B and C in Eq. (66) are quite different, and the Born plane wave trial functions apparently are not close to the correct wave functions. To make B and C equal it is necessary to introduce a considerable modification of the trial function from the Born plane waves. On the other hand, such a modification in the direction of making B and C equal generally will change the magnitude of B, which in the present Yukawa case pulls the variational estimate of the differential cross section away from the very good Born value. To sum it up, when the magnitude of the second Born correction term is large, we may expect to have to use a considerably different trial function than the Born plane wave in order to get a reasonably good estimate of the differential cross section. Under such circumstances, it is difficult to get good results with the variational principle since Born plane waves so often provide, as in the calculations we are reporting, the only reasonable and tractable starting point.

TABLE II. Differential cross sections at 20 Mev.

	Differential cross sections Variational formulation III Even waves: $K=K_s$ Odd waves: K equal to		
Angle	K_a	K_b	Exact
0	4.45	4.28	3.99
$\pi/2$	2.26	2.26	2.28
π	1.92	1.95	2.27

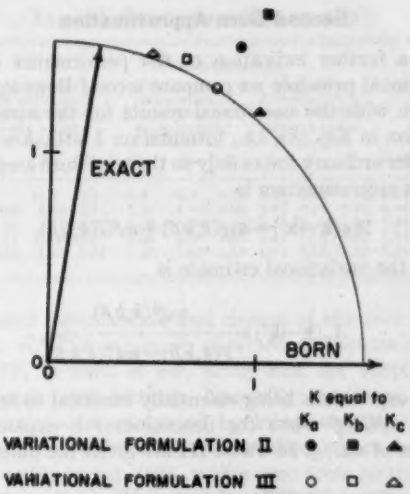

FIG. 8. Argand diagram for the even complex scattering amplitude at 20 Mev in the forward direction. The arc is drawn with radius equal to the magnitude of the exact amplitude.

If the imaginary part of C is small for Born trial functions, B and C may be made fairly close by introducing modifications to the trial function which keep B real. Moreover, since the differential cross section adds the real and imaginary parts in quadrature, it is clear that we can get a quite good variational estimate of the differential cross section in this way. When the imaginary part of C is appreciable however, we must somehow modify the trial function so that B is complex. In the Yukawa case we have been studying, the imaginary part of C which equals the imaginary part of $u_0^3 G(k,k,\theta)$ actually is large. Thus we see one obvious reason why the modification provided by formulation II, which keeps B real whatever the choice of K, is inferior to the more general projection operator formulation III, in which the imaginary factors $e^{\pm 2\pi i/3}$ explicitly guarantee that B will be complex. As a matter of fact comparison of our numerical values in this formulation shows that the complex quantities B and C do turn out reasonably close to each other in formulation III. This suggests that for a potential for which exact differential cross sections are not available, the difference between B and C may prove useful as a measure of the rate of convergence of the variational method to the correct values.[20]

[20] Since B is proportional to the amplitude α of the trial function, and C is proportional to α^3, it is clear that $B-C$ or the relative difference $(B-C)/C$ may be made as large or small as desired merely by an appropriate choice of α. Thus, with a sequence of trial functions this measure of the rate of convergence is meaningful only if the amplitudes of the functions are maintained relatively constant, by fixing the value at the origin for example, or better perhaps by fixing $\int d\mathbf{x}\psi^* u\psi$.

Second Born Approximation

As a further indication of the performance of the variational principle we compare second Born approximation with the variational results for the same trial function in Eq. (8), i.e., formulation I with $K=k$. We consider ordinary forces only so that the Born amplitude in this approximation is

$$A_B(\mathbf{k}\to\mathbf{k}')=u_0g(k,k,\theta)+u_0^2G(k,k,\theta), \tag{67}$$

while the variational estimate is

$$A_v(\mathbf{k}\to\mathbf{k}')=\frac{u_0g^2(k,k,\theta)}{g(k,k,\theta)-u_0G(k,k,\theta)}, \tag{68}$$

these expressions being essentially identical to terms of order $u_0^3G^2/g^2$ when the denominator is expanded in powers of u_0G/g. In Table III are given the differential

TABLE III. Differential cross sections in second Born approximations.

Energy (Mev)	Angle	Differential cross sections: Exact	Variational formulation I with $K=k$	Second Born approximation	Expansion parameter u_0G/g
20	0	3.99	8.61	13.2	$0.43+0.57i$
	$\pi/2$	2.28	1.51	6.1	$0.69+0.98i$
	π	2.27	0.41	4.03	$0.89+1.33i$
90	0	5.07	6.27	8.0	$0.13+0.37i$
	$\pi/2$	0.309	0.121	0.74	$0.34+1.20i$
	π	0.151	0.023	0.336	$0.42+1.66i$
150	0	5.31	6.00	7.1	$0.08+0.30i$
	$\pi/2$	0.127	0.048	0.30	$0.25+1.21i$
	π	0.048	0.009	0.125	$0.29+1.64i$

cross sections computed from Eqs. (67) and (68), at 20, 90, and 150 Mev, at angles of 0°, 90°, 180°, with again the exact values for comparison. The variational entries are of course identical with those in the $K=K_a$ variational formulation I column of Table I. We also tabulate the expansion parameter u_0G/g which, for Born trial functions, is the fractional difference between B and C and is therefore a reasonable measure of the accuracy of the variational method, as well as of the convergence of the Born expansion. Evidently the variational estimates are better than the second Born except at those energies and angles where the expansion parameter is so large that both the Born and variational estimates are extremely poor and their comparison is essentially meaningless. The variational phase angles are similarly closer to the exact phases than are the second Born approximation phases. With an arbitrary potential, where the first Born approximation generally is not close to the magnitude of the correct scattering amplitude, this comparison of the second Born approximation with the variational principle for the same trial function would be expected to be more indicative of the performance and utility of the variational method than comparison with first Born approximation, for moderate values of the expansion parameter. The results in Table III for the present Yukawa case suggest that with other potentials the variational method generally will indeed be an improvement on second Born approximation, and *a fortiori* on first Born approximation, even without improvement of the trial functions from first Born plane waves.

High-Energy Limit

In the foregoing discussion we have been thinking mainly of the intermediate energy region below those energies at which the usual criteria for the validity of the Born approximation are satisfied. For any presently reasonable potential purporting to describe neutron-proton scattering this region will be roughly the same 20–150 Mev region we have been considering. As the high-energy Born limit is approached with such a potential, we might expect the second Born correction term to become negligible.[21] Similarly the exact scattering phases ought to approach the Born value of zero, a behavior illustrated in Fig. 4. Simultaneously, according to the remarks in the preceding paragraph, the variational method even with Born trial functions, should generally be better than the Born and become more and more accurate, except that at high enough energies both the Born and the variational principle should yield results indistinguishable from the exact values.

For most potentials these expectations are fulfilled at zero scattering angle, as we can demonstrate. We employ the cross-section theorem,

$$\sigma=\frac{4\pi}{k}\,\mathrm{Im}A(\mathbf{k}\to\mathbf{k})=\frac{4\pi}{k}\,\mathrm{Im}A(0), \tag{69}$$

where σ is the total cross section, and $A(0)$ is the forward scattering amplitude. In the high-energy limit, $\sigma\simeq\pi\theta^2|A(0)|^2$, where $\theta\simeq 1/k$ is the width of the diffraction peak, with k in units of the range of the potential. Then from Eq. (69) we obtain

$$\frac{\mathrm{Im}A(0)}{|A(0)|}\simeq\frac{1}{4k}|A(0)|, \tag{70}$$

which approaches zero as $k\to\infty$, since $|A(0)|$ is quite generally bounded.

At angles for which the differential cross section is not a violently fluctuating function of angle, it is not difficult to prove, using an extension of the cross section theorem, that the phase still tends to zero.[22] If the

[21] In fact for our Yukawa case, examination of the high-energy limit of Eqs. (60) and (62) shows that Eq. (65) guarantees $u_0G/g\ll 1$, except that outside of the forward peak, $k\sin\theta/2\gg 1$, we require the somewhat stronger condition $(u_0/k)\log[4k^2\sin^2(\theta/2)]\ll 1$ to make the imaginary part of the ratio small. This extra condition probably is connected with the Coulomb-like nature of the scattering at wide angles or small impact parameters.

[22] For the Yukawa well this prediction seems to be borne out by our results.

cross section does fluctuate violently, however, it appears that large phase shifts may be expected. Thus we may draw a distinction between the high-energy limits at angles within the main diffraction peak, $k\sin(\theta/2)\ll 1$, and angles outside, $k\sin(\theta/2)\gg 1$. Within the peak, for most potentials, the cross section varies smoothly and the phase approaches zero in the Born limit. Outside the peak, even though the usual criteria for the validity of Born approximation may be satisfied, the phases in the high-energy limit generally are zero only for those potentials for which the true cross sections at those angles do not fluctuate markedly. Moreover, when such fluctuations occur, the first Born approximation cross sections can be expected to exhibit similar but certainly not coincident fluctuations, showing that the magnitude as well as the phase of the scattering amplitude has been poorly estimated. Correspondingly, we anticipate that at high energies the variational principle will be quite generally useful within the main diffraction peak, but that its utility outside the peak will depend on the potential. Where the cross sections fluctuate violently, the variational method probably will readily account for the qualitative wide angle behavior of the differential cross section, but will converge only slowly, i.e., will require trial functions very different from Born plane waves in order that further improvements in the trial functions change the results only slightly. Hence for such potentials it appears likely that at wide angles the variational method will be most useful and accurate at intermediate energies.

Total Cross Sections

We conclude by describing our results for the total cross sections. These have been computed by the variational method using Eq. (69), for the various energies, variational formulations, and choices of effective wave number K. The values so obtained are tabulated in Table IV, in units of πa^2, along with the exact and Born total cross sections. Naturally the Born total cross sections are computed by integrating the Born approximation differential cross sections, since for the Born amplitude Eq. (69) implies zero cross section. It is seen from Table IV that the variational formulations give fair estimates of the total cross section in almost all cases. These results are about as accurate as those reported earlier for the square well.[7] It is also seen from Table IV that, surprisingly enough, formulation II is usually superior to formulation III. We have no very good explanation of this unexpected result. As is apparent from inspection of the figures, however, total cross sections obtained by integration of the differential cross sections determined by variational principle III would give results in good agreement with the exact values at all energies and better than either formulation II or the Born total cross section.

TABLE IV. Total cross sections vs energy.

	Total cross section										
	Variational formulation I with effective wave number K equal to			Variational formulation II with effective wave number K equal to			Variational formulation III with effective wave number K equal to				
Energy (Mev)	K_a	K_b	K_c	K_a	K_b	K_c	K_a	K_b	K_c	Born	Exact
20	12.5	11.7	4.31	10.3	10.8	6.23	8.86	9.44	9.08	8.12	10.1
50	5.05	4.90	2.82	4.61	4.64	3.05	3.82	3.99	3.54	4.16	4.53
90	2.82	2.77	1.94	2.70	2.69	1.98	2.33	2.38	2.06	2.51	2.64
120	2.11	2.09	1.57	2.06	2.05	1.59	1.83	1.86	1.61	1.94	2.01
150	1.69	1.67	1.33	1.66	1.65	1.33	1.51	1.53	1.33	1.58	1.63

We should like to express our appreciation to Miss Patricia Childress of the National Bureau of Standards who performed most of the numerical computations reported above.

A Possible Relation between the Pseudoscalar β Coupling and the Nuclear Spin-Orbit Coupling*

E. J. KONOPINSKI
Indiana University, Bloomington, Indiana
(Received January 13, 1954)

It is suggested that there is a direct proportionality between the nuclear spin-orbit energy and the size of the nuclear moments responsible for β decay via pseudoscalar coupling. On this assumption, the observed size of the spin orbit splitting is adequate to account for the large β moment needed for explaining the RaE spectrum, within the uncertainties in the knowledge of these phenomena.

1.

PREVIOUS evaluations[1] of the pseudoscalar (P) β coupling, if correct, predict for it a negligible effectiveness: the corresponding nuclear moment is saved from vanishing only by relativistic effects, inversely proportional to the large nucleon mass. On the other hand, an anomaly of long standing, the spectrum of RaE, appears to require for its explanation that the P coupling contribution be orders of magnitude larger.

A similar contradiction of long standing exists in connection with the well-known spin-orbit couplings in nuclei. A large spin-orbit force seems essential for understanding the shell model. Moreover, its presence seems to be quite directly confirmed by the recent[2] measurements on the polarization of neutrons scattered by nuclei. A natural way[3] for a spin-orbit force to arise, is as a relativistic correction to the main attractive potential. Again, this is an effect inversely proportional to the nucleon mass (squared) and expected to be much too small to account for the above observations.

It will be pointed out here that there exists a possible basis for a close relationship between the two contradictions, one which seems to have escaped notice so far. On the basis to be suggested, both the RaE spectrum, and the spin-orbit force evidence, become measures of a single quantity: ξ, if $-\xi\boldsymbol{\sigma}\cdot\mathbf{L}$ is the spin-orbit energy.

The point to be made here may be viewed as follows: The above-mentioned evaluations of the pseudoscalar β moment rely on certain current phenomenological ideas concerning the nuclear Hamiltonian. However, the similarly phenomenological spin-orbit force has been ignored in those considerations. It will be assumed here that just this large spin-orbit energy magnitude furnishes the essential information for obtaining the β moment. The latter then attains a magnitude adequate for explaining the RaE spectrum.

A sufficiently large β moment has already been obtained from another point of view by Ruderman.[4] He goes back to the fundamental pion-field origins of the nuclear forces. The results are beset by the usual uncertainties concerning the adequacy of first-order perturbation theory. Moreover, the same point of view has not so far accounted for the simpler phenomenologically established features of the nuclear Hamiltonian. Finally, of course, one expects to bring about complete consistency between the phenomenological picture and the meson field basis. The point to be made here indeed suggests that the explanation of the spin-orbit coupling should be looked for in the same basic mesonic processes which are held to account for the pseudoscalar β moment.

2.

The most attractively simple attempt to account for the spin-orbit forces was made by Furry.[3] He assumed that the attractive nuclear potential, $-U$, fits into the Dirac Hamiltonian for a nucleon as a relativistic invariant:[5]

$$H=\boldsymbol{\alpha}\cdot\mathbf{p}+\beta(M-U). \tag{1}$$

It follows directly that the "small" component of the energy eigenfunction is

$$\chi=f\boldsymbol{\sigma}\cdot\mathbf{p}\Psi/2M, \quad f=[1+(E-U)/2M]^{-1}\approx 1, \tag{2}$$

where Ψ is the Pauli wave function ("large" component) and $E+M$ is written for the energy eigenvalue.

The nonrelativistic Hamiltonian $E=H-M$ becomes

$$E=fp^2/2M-U-i\nabla f\cdot\mathbf{p}/2M+\boldsymbol{\sigma}\cdot[\nabla f\times\mathbf{p}]/2M. \tag{3}$$

The last, spin-dependent, term in the energy becomes

$$E_s=\boldsymbol{\sigma}\cdot[\nabla U\times\mathbf{p}]/4M^2f^2. \tag{4}$$

One treats the nuclear potential as effectively central, $U(r)$, in order to have conservation of the individual nucleon's total angular momentum, as required by the shell model. Now $\nabla U=(U'/r)\mathbf{r}$ and

$$E_s=-\xi\boldsymbol{\sigma}\cdot\mathbf{L}, \quad \text{with} \quad \xi=-U'(r)/4M^2f^2r. \tag{5}$$

This shows the advantage of Furry's introduction of the nuclear potential as a relativistic scalar. It has led to a sign of E_s opposite to what is obtained when the potential is treated as the fourth component of a four-

* Supported by a grant from the National Science Foundation.

[1] Ahrens, Feenberg, and Primakoff, Phys. Rev. 87, 663 (1952); T. Ahrens, Phys. Rev. 90, 974 (1953); M. E. Rose and H. K. Osborn, Phys. Rev. 93, 1315 (1954).

[2] H. Barschall, Phys. Rev. 93, 917 (1954).

[3] D. R. Inglis, Phys. Rev. 50, 783 (1936); W. H. Furry, Phys. Rev. 50, 784 (1936); G. Breit, Phys. Rev. 51, 248 (1937).

[4] M. Ruderman, Phys. Rev. 89, 1227 (1953).

[5] We use units $\hbar=c=m=1$ except where otherwise specified.

vector. The latter must be done for atomic electrons, which are subject to electrostatic force. In this way the inversion[5] of the nuclear spin-orbit doublets relative to atomic ones emerges automatically.

The spin-orbit energy coefficient ξ is usually[3] estimated as

$$\xi_{\text{theor}} \approx \bar{U}/4M^2R^2 \approx (1/11A^{\frac{1}{3}}) \text{ Mev}, \tag{6}$$

where $R=1.4A^{\frac{1}{3}}\times 10^{-13}$ cm is the nuclear radius, and $\bar{U}$ is taken to be a nuclear potential depth ≈ 20 Mev. Of course, the estimate $\bar{U}'\approx -\bar{U}/R$ is very questionable in view of the possible singularity of the force at an individual nucleon. It seems preferable to forego theoretical evaluation at this stage and use instead the figure roughly indicated by the evidence[3]

$$\xi_{\text{exp}} \approx 1-2 \text{ Mev} \tag{7}$$

for $A\approx 100$, even though this is 200 to 400 times larger than (6). Of course, this procedure is subject to the uncertainty of whether the spin-orbit effect has been traced to its true origin.

3.

We turn now to the problem of RaE. According to Petschek and Marshak's analysis,[6] the spectrum is a resultant of contributions from both tensor (T) and pseudoscalar β couplings. The T interaction works through the nuclear moment $\langle\boldsymbol{\sigma}\cdot\mathbf{r}\rangle$, while the pseudoscalar contribution is determined by the size of $\langle i\beta\gamma_5\rangle$. The observed spectrum shape is reproduced if the ratio,

$$\Gamma = -\frac{G_P}{G_T}\frac{\langle i\beta\gamma_5\rangle}{\langle\boldsymbol{\sigma}\cdot\mathbf{r}\rangle} \approx 13, \tag{8}$$

is adopted. Here $\mathbf{r}$ is in units of $\hbar/mc$. G_P and G_T are the pseudoscalar and tensor coupling constants respectively. G_P is unknown but cannot reasonably be expected to be larger than G_T by more than a factor 3 or so.[7]

For the nuclear moments, it is consistent with the shell model to use the single-nucleon state with respect to which RaE and its daughter nucleus differ ($\psi_i \rightarrow \psi_f$). Then

$$\langle i\beta\gamma_5\rangle \equiv \int d\tau \psi_f^\dagger i\beta\gamma_5\psi_i = -i\int d\tau \Psi_f^\dagger \chi_i + i\int d\tau \chi_f^\dagger \Psi_i, \tag{9}$$

where, as in Sec. 2, Ψ and χ refer to the "large" and "small" components, respectively. According to (2),

$$\begin{aligned}\int d\tau \chi_f^\dagger \Psi_i &= \int d\tau (f\boldsymbol{\sigma}\cdot\mathbf{p}\Psi_f)^\dagger \Psi_i/2M \\ &= \int d\tau \Psi_f^\dagger \boldsymbol{\sigma}\cdot\mathbf{p} f\Psi_i/2M \\ &= \int d\tau \Psi_f^\dagger \chi_i - (i/2M)\int d\tau \Psi_f^\dagger \boldsymbol{\sigma}\cdot(\nabla f)\Psi_i. \end{aligned} \tag{10}$$

Thus,

$$\begin{aligned}\langle i\beta\gamma_5\rangle &= \int d\tau \Psi_f^\dagger (\boldsymbol{\sigma}\cdot\nabla U/4M^2 f^2)\Psi_i \\ &= \int d\tau \Psi_f^\dagger (U'/4M^2 f^2 r)\boldsymbol{\sigma}\cdot\mathbf{r}\Psi_i \\ &\approx -(\bar{U}/4M^2R^2)\langle\boldsymbol{\sigma}\cdot\mathbf{r}\rangle, \end{aligned} \tag{11}$$

under the same assumptions as those leading to (5) and (6). Within the accuracy of these considerations the coefficient here is the same as ξ of (6), hence

$$\Gamma = \xi G_P/G_T. \tag{12}$$

This demonstrates the possible close relationship between the β spectrum datum, Γ, and the spin-orbit coefficient, ξ.

If one now takes the "theoretical" estimate (6) for ξ, one predicts

$$\Gamma \approx (1/175)G_P/G_T, \tag{13}$$

which, as in former evaluations, is much too small to account for the RaE spectrum. However, once the basis of comparison (12) is accepted, it becomes more reasonable to take $\xi_{\text{exp}} \approx (2 \text{ to } 4)mc^2$ as in (7). Now,

$$\Gamma \approx (2 \text{ to } 4)G_P/G_T, \tag{14}$$

which can no longer be said to disagree in order of magnitude with Marshak and Petschek's value $\Gamma\approx 13$, considering the many uncertainties involved (e.g., the unknown value of the spin-orbit coefficient in the neighborhood of $A\approx 200$; the finite de Broglie wavelength effect on Γ; etc.).

We note here the relationship of those considerations to such as those of Ahrens and Rose.[1] The contribution of the P coupling to β decay is more properly determined by the nuclear matrix element $\langle i\beta\gamma_5 B\rangle$ where $B=\psi_\nu^\dagger i\beta\gamma_5\psi_e$ is the electron-neutrino covariant. Then (9), (10), and (11) are replaced by

$$\langle i\beta\gamma_5 B\rangle = \int d\tau \Psi_f^\dagger \boldsymbol{\sigma}\cdot\nabla(fB)\Psi_i/2M. \tag{15}$$

Ahrens and Rose consider the variation of B over the nucleus as the important effect, hence even obtain a somewhat differently shaped spectrum than that used by Marshak and Petschek. If this effect is the main one, then there is no basis (such as suggested here) for supposing the pseudoscalar β coupling to be other than negligible. The variation of the lepton waves (i.e., B) over the nucleus is just the phenomenon responsible for "forbiddenness" and puts the contribution into the "third-forbidden" class of magnitudes. The suggestion of the present work is that, instead, the $f\nabla B$ term of $\nabla(fB)$ in (15), is negligible compared to $B\nabla f$.

[6] A. G. Petschek and R. E. Marshak, Phys. Rev. 85, 698 (1952).

[7] $G_P=+3G_T$ would fit the latest measurements of the μ-decay spectrum, if only a scalar (S) contribution may exist, in addition, as required by the β-decay data.

Letters to the Editor

PUBLICATION of brief reports of important discoveries in physics may be secured by addressing them to this department. The closing date for this department is five weeks prior to the date of issue. No proof will be sent to the authors. The Board of Editors does not hold itself responsible for the opinions expressed by the correspondents. Communications should not exceed 600 words in length and should be submitted in duplicate.

Luminescence of Beryllium Oxide*

C. E. MANDEVILLE AND H. O. ALBRECHT
Bartol Research Foundation of The Franklin Institute, Swarthmore, Pennsylvania
(Received February 15, 1954)

IN the course of a survey of the luminescence of metallic oxides, some interesting properties of beryllium oxide have been noted. When irradiated by alpha particles, an intense ultraviolet fluorescence was observed. From the spectral sensitivity of the detector, it was estimated that a considerable emission occurs at wavelengths of 2500A or less.

Although earlier investigators have reported a visible blue luminescence from BeO under action of a vacuum discharge, apparently no mention has been made of the fact that the spectrum extends far into the ultraviolet. Under alpha irradiation, white powdered BeO yielded an emission band extending from above 4000A into the ultraviolet to ~2000A. The spectrogram was obtained in a small Hilger quartz spectrograph. Bombardment of a combustion tube of sintered BeO gave the same result. BeO has been reacted with other materials, such as SiO_2, to produce numerous phosphors but does not itself appear to have been studied in detail. A comparison of the intensity of the ultraviolet fluorescence of powdered BeO and of NaCl—Ag under alpha-particle excitation is shown in Table I.

Nine different samples of BeO were obtained from eight different suppliers, and all exhibited a detectable long-period phosphorescence as well as energy storage at room temperature (20°C) when excited by cathode rays, x-rays, or ultraviolet light. Irradiation by alpha particles resulted principally in fluorescence. One sample of BeO, excited by cathode rays, decayed over a period of 400 minutes on a $\log I - \log t$ plot with a slope of -0.83. The decay of both visible and ultraviolet portions of the spectrum followed the same law. Spectrographic analyses showed that B, Mg, Al, Si, and Fe could be present as impurities. The afterglow which immediately followed excitation could be considerably enhanced in the case of the less pure samples by ignition at 1400°C. Under cathode-ray excitation, this afterglow equaled in intensity that of the best samples of NaCl—Ag. The purest material was obtained from Johnson, Matthey, and Company, Ltd., and although it had good fluorescent response for alpha and beta particles, phosphoresced only faintly under irradiation by excitants of all types. Ignition of this particular material had small effect upon the phosphorescence and energy storage. The alpha-particle-excited fluorescence of all samples seemed to remain unchanged by ignition.

TABLE I. Ultraviolet fluorescence.

Phosphor*	Yield
NaCl+0.0008 percent AgCl	31
NaCl+0.004 percent AgCl	71
NaCl+0.02 percent AgCl	370
NaCl+0.10 percent AgCl	1160
NaCl+0.22 percent AgCl	3150
NaCl+0.50 percent AgCl	4630
NaCl+1.00 percent AgCl	3730
BeO	1060
BeO (ignited)	990

* AgCl concentrations in weight percent.

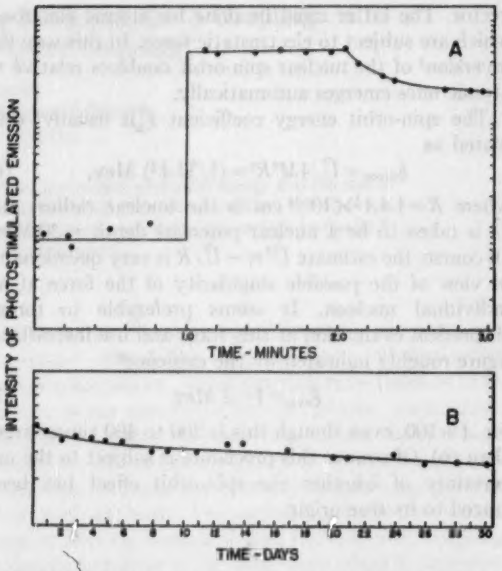

FIG. 1. Photostimulated emission of x-rayed beryllium oxide. In Fig. 1(A), photostimulated emission occurs during one minute of stimulation and is followed by a post-stimulation afterglow. Figure 1(B) is a plot of the decay of the energy stored in x-rayed BeO. Excitation and all measurements carried out at room temperature (20°C).

A sample of BeO obtained from Fisher Scientific Company, Inc., was ignited at 1400°C and irradiated by 20-kev x-rays and stored in darkness. After twenty-two hours, the excited phosphor was photostimulated by light from a tungsten lamp for one minute. The emission during stimulation and the post-stimulation afterglow are shown in Fig. 1(A). A similar specimen was x-rayed, and bursts of ultraviolet of duration six seconds were drawn from it by photostimulation at intervals of 24 to 48 hours over a period of thirty days. This photostimulated emission is approximately proportional to the amount of energy remaining stored in the beryllium oxide. These data are plotted in Fig. 1(B), where it is clear that a large fraction of the stored energy remained after thirty days—more than in any other phosphor studied in this laboratory. Under identical conditions of photostimulation, the emission from BeO was less than one percent of that from transparent samples of NaCl—Ag of the same area. However, the transmission of the powdered BeO is low, and only light released near the surface could be measured.

* Assisted by the joint program of the U. S. Office of Naval Research and the U. S. Atomic Energy Commission.

Restoration of F and V_1 Centers Following Low-Temperature Bleaching

R. V. HESKETH AND E. E. SCHNEIDER*
King's College, University of Durham, Newcastle upon Tyne, England
(Received February 24, 1954)

THE following observations were made in the course of experiments with single crystals of KCl colored at 113°K by 180-kv x-rays.

The freshly colored crystals contain a large fraction (up to 50 percent) of unstable F centers which require a thermal activation of only 0.02 ev in order to decay, presumably by tunneling of electrons to neighboring positive hole centers. This is revealed by

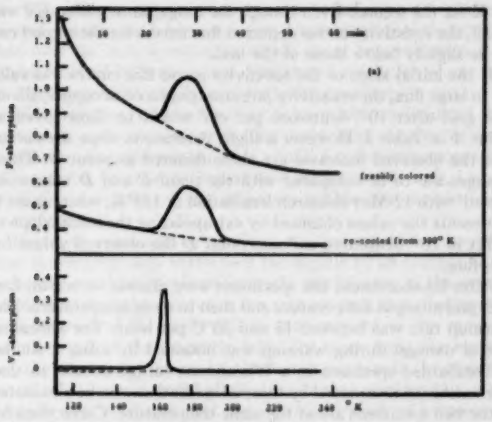

FIG. 1. F and V_1 absorption during warming after optical bleaching at 113°K.

a steep drop in the "F-absorption glow curve," i.e., the curve giving the absorption in the maximum of the F band as a function of time or temperature during gradual warming of the crystals. The unstable centers are removed by raising the crystal to room temperature so that the absorption glow curve obtained after recooling to 113°K shows only a slight fall entirely explained by the broadening and shift of the maximum of the F band with rising temperature.

Optical bleaching by F irradiation occurs at 113°K with a saturation value of 30 percent of the F centers bleached, and is followed by a partial recovery in the dark in which up to $\frac{1}{3}$ of the bleached F centers reappear within about 10 minutes. The effects are similar to those reported by Markham, Platt, and Mador[1] on x-rayed KBr at 78°K. But their explanation of the recovery in terms of a dissociation of F' centers cannot apply in our case where F irradiation has been found to reduce and not to enhance the F' band. Instead, we believe that the recovery is connected with a narrow (0.17 ev) absorption band at 8750A which we have found to appear during optical F bleaching.

The result of main interest concerns a large-scale restoration phenomenon which occurs if the crystals are first bleached at 113°K by F irradiation and then slowly warmed in the dark (2.3°K per min). The F-absorption glow curves observed under these conditions for a virgin crystal and a crystal recooled from room temperature (Fig. 1(a)) show a large peak around 175°K superimposed on the normal absorption glow curves discussed above. This means that F centers are temporarily recreated and subsequently bleached thermally. The thermal bleaching can be prevented and the restored absorption be arrested by rapid recooling from a temperature near the restoration peak (Fig. 2). It is seen that nearly all the optically bleached F centers can be regained by this procedure.

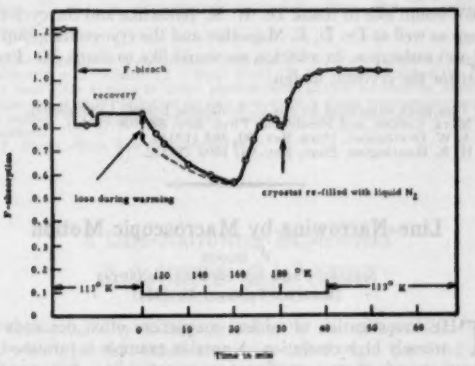

FIG. 2. Restoration of optically bleached F centers by warming and recooling.

Even more remarkable is the fact that simultaneously with the F centers, V_1 centers are recreated as revealed by a sharp peak in the absorption glow curve obtained by observing the absorption in the maximum of the V_1 band during warming. The curve shown in Fig. 1(b) refers to a crystal recooled from room temperature so that at the beginning of the experiment it did not contain any V_1 centers but only more stable higher V centers. The fact that the peak occurs slightly below that for the F centers and is very much sharper implies a very short lifetime of the restored V_1 centers in accordance with their low dissociation temperature of 128°K as observed by Dutton and Maurer.[2] It appears very significant that the height of the V_1 restoration peak, representing only a fraction of the total number of transient V_1 centers involved, is about twice as large as the height of the original V_1 band present in the crystals immediately after x-raying.

In explanation of the effects described we offer the following tentative suggestions: Crystals x-rayed at low temperature contain a large number of positive holes in so far optically unidentified V_0 centers. Electrons released by F irradiation are trapped at these centers to form complex metastable centers (possibly associated with the peak at 8750A) which on breakup by thermal activation yield unstable F centers and V_1 centers in close proximity to each other.

* 1953–1954 at Duke University, Durham, North Carolina.
[1] Markham, Platt, and Mador, Phys. Rev. 92, 597 (1933).
[2] D. Dutton and R. Maurer, Phys. Rev. 90, 126 (1953).

Effect of Hydrostatic Pressure on the Superconducting Transition of Tin and Thallium

MILAN D. FISKE
General Electric Research Laboratory, Schenectady, New York
(Received February 19, 1954)

THE dependence on pressure of the magnetic threshold field of superconductivity has been measured for tin and for thallium for pressures transmittable by liquid helium. The experimental technique consisted of measuring the difference of the threshold fields of two similar ellipsoidal single crystals at the same temperature. The pressure on one ellipsoid was held fixed while the pressure on the other was varied by means of tank helium gas. The transition was followed by means of 60-cycle ac susceptibility measurements.

The summary of results to date on tin and thallium is given in Table I, which contains a review of previous determinations of the pressure coefficients. The values for tin marked by asterisks are deduced from data on wires in tension, in which the hydrostatic stress is one-third the tensile stress for isotropic materials. The rather good agreement among the tin data between the pressure coefficients directly measured in hydrostatic experiments and those deduced from wire straining suggests that in the latter the shear components have a small or negligible effect on the superconducting transition in tin. This speculation is now being checked with a thin-walled tin cylinder stressed in pure shear.

No ready explanation is at hand for reconciling the results of the present work on thallium and those of Kan, Lazarev, and Sudovstov[1] with those recently reported by Chester and Jones,[2] differing as they do in both magnitude and sign. The technique employed by Chester and Jones results in massive plastic deformation at room temperature while that used by Kan *et al.* may produce some plastic deformation near the ice point. To examine the effects of mild plastic deformation, an ellipsoid was machined from a single crystal twisted 180° over the 1.2-cm length of the ellipsoid. The threshold curve and pressure coefficient for this sample were the same as for the undeformed crystal.

Bridgman's phase diagram[3] shows that only α thallium (h.c.p.) should be expected in these experiments. However, quenched β thallium (f.c.c.) is stable at room temperature.[4] An elongated specimen selected from particles produced by pouring molten thallium into liquid nitrogen had the same characteristics as previous samples.

These measurements, which will be reported in detail later, are being continued on specimens for which the crystal structure is known, using apparatus permitting appreciable increased accuracy of determination of the pressure coefficients at zero pressure.

TABLE I. Summary of data on the effect of hydrostatic pressure on the superconducting transition of tin and thallium.[a]

Authors	Pressure (atmos)	ΔT_c (°K)	$(\Delta T_c/\Delta P)\times 10^4$ deg atmos^{-1}
	For tin		
Sizoo and Onnes[b]	95	−0.0027	−3
	82[c]	−0.007	−8.6[c]
Alekseyevsky[d]	21[c]	−0.001	−5
Kan, Lazarev, and Sudovstov[e]	1730	−0.10	−5.8
Chester and Jones[f]	11 500	−0.52	−4.3
	17 500	−0.8	−4.5
Grenier, Spöndlin, and Squire[g]	60[c]	−0.0034	−5.5[c]
Present work	<115		−4.9±0.5
	For thallium		
Kan, Lazarev, and Sudovstov[h]	1370	0.008	0.6
	1730	0.025	1.4
Chester and Jones[f]	13 400	−0.06	−0.4
Present work	<48		1.3±0.2

[a] Presented at the Third International Conference on Low Temperature Physics and Chemistry, Houston, Texas, December 17, 1953.
[b] G. J. Sizoo and K. Onnes, Leiden Comm. No. 180b (1925).
[c] Hydrostatic value deduced from tensile stress.
[d] N. E. Alekseyevsky, J. Exptl. Theoret. Phys. (U.S.S.R.) 10, 746 (1940).
[e] Kan, Lazarev, and Sudovstov, Doklady Akad. Nauk. 69, 173 (1949).
[f] See reference 2.
[g] Grenier, Spöndlin, and Squire, Physica 19, 833 (1953).
[h] See reference 1.

[1] Kan, Lazarev, and Sudovstov, J. Exptl. Theoret. Phys. (U.S.S.R.) 18, 825 (1948).
[2] P. F. Chester and G. O. Jones, Phil. Mag. 44, 1281 (1953).
[3] P. W. Bridgman, Phys. Rev. 48, 893 (1935).
[4] S. Sekito, Z. Krist. 74, 189 (1930).

Resistivity Changes in Copper, Silver, and Gold Produced by Deuteron Irradiation Near 10°K*

H. G. COOPER, J. S. KOEHLER, AND J. W. MARX
Department of Physics, University of Illinois, Urbana, Illinois
(Received February 19, 1954)

PREVIOUS research[1,2] on various metals indicates that radiation damage introduced near liquid nitrogen temperature will anneal when held at the irradiation temperature. The present experiment was designed to obtain damaged specimens in which no thermally activated motion of the lattice defects occurs during irradiation.

99.97±0.02 percent pure wires of copper, silver, and gold were irradiated with 12-Mev deuterons. The wire diameters were 5 mils for the copper and silver and 4 mils for the gold. The maximum temperature during bombardment was 12°K for the first run and 16°K for the second. Even though the temperature difference was small, the resistivity *versus* deuteron flux curves for the second run all lie slightly below those of the first.

TABLE I. Radiation-induced resistivity increases, in units of 10^{-7} ohm cm, after a flux of 10^{17} deuterons/cm^2.

Point	Copper	Silver	Gold
A extrapolating initial 12°K slope	2.3	2.6	3.8
B (observed 12°K value)	1.9	2.0	3.1
C extrapolating initial 135°K slope	0.9	1.4	2.6
D (observed 135°K value)	0.6	1.0	1.9

If the initial slope of the resistivity *versus* flux curves was valid up to large flux, the resistivity increases produced in copper, silver, and gold after 10^{17} deuterons per cm^2 would be those given as point *A* in Table I. However a slight decrease in slope does occur and the observed increases are those denoted as point *B*. These changes are to be compared with the point *C* and *D* values obtained[1] with 12-Mev deuteron irradiation at 135°K, where point *C* represents the values obtained by extrapolating the initial slope to a flux of 10^{17} deuterons/cm^2 and point *D* the observed values for this flux.

After bombardment the specimens were allowed to warm, first to liquid nitrogen temperature and then to room temperature. The warmup rate was between 15 and 30°C per hour. The annealing out of damage during warmup was obtained by using a similar unbombarded specimen in a Wheatstone bridge circuit[2] so that the resistivity introduced by thermal oscillations can be eliminated if the two specimens are at the same temperature. Corrections for temperature differences of specimen and dummy were obtained by comparison with warmups made before irradiation and after annealing to room temperature.

On holding for 48 hours at or slightly below the bombardment temperature, no change in resistivity (i.e., to ±0.1 percent) was observed. In copper and silver, abrupt drops in resistivity indicated that some rather unique annealing process occurred at about 40°K and 30°K, respectively. In gold no large low-temperature drops were found. The first two rows in Table II indicate the magnitude of these processes and the third gives the percentage left after annealing to liquid nitrogen. Two possible processes that may explain these abrupt recoveries are the motion of interstitial atoms, which Huntington[3] calculated would have an activation energy in the range 0.07 to 0.24 ev, or the recombination of very close interstitial-vacancy pairs as suggested previously.[1,2]

TABLE II. Percentage of initial resistivity increase remaining after annealing to various temperatures.

Annealing temperature	Copper	Silver	Gold
35°K	90	78	97
45°K	50	77	93
77°K	41	69	86
220°K	25	32	55
300°K	7	10	10

A gradual decrease occurs in all three materials during the warmup to 220°K and then a more rapid process becomes important.[1,2] The latter process is complete at about 255°K in copper, at approximately 240°K in silver, and at roughly 285°K in gold. The percentage remaining at both 220°K and 300°K is given in Table II.

We would like to thank Dr. W. K. Jentschke and the cyclotron group as well as Dr. D. E. Mapother and the cryogenics group for aid and assistance. In addition we would like to thank Mr. Frank Witt for the cryostat design.

* This work supported by the U. S. Atomic Energy Commission.
[1] Marx, Cooper, and Henderson, Phys. Rev. 88, 106 (1952).
[2] A. W. Overhauser, Phys. Rev. 90, 393 (1953).
[3] H. B. Huntington, Phys. Rev. 91, 1092 (1953).

Line-Narrowing by Macroscopic Motion

F. BLOCH
Stanford University, Stanford, California
(Received February 19, 1954)

THE investigation of nuclear magnetism often demands extremely high resolution. A notable example is furnished by the structure in the groups of proton resonance lines discovered by Arnold and Packard.[1] The separation of the three components of

the CH_3 group in ethyl alcohol amounts to only about 6 cycles per second and requires for its resolution a homogeneity of the external field over the sample region to within about one milligauss. In a typical field of 7000 gauss this represents a resolution of one part in 7 millions which can be achieved with considerable care.

It is tempting, however, still to increase the resolution at least tenfold in order to observe further interesting details. While it would seem very difficult to obtain a field of sufficient homogeneity, it is possible to reach this goal in a rather simple way, i.e., by providing a motion of the substance within the sample under investigation.

The mechanism can be qualitatively understood by assuming that the external field varies over the sample by an amount of the order ΔH and that a given molecule participates in the macroscopic motion of the substance in a manner such that, during each time interval of the order t, it is once exposed to the range of variation ΔH of the magnetic field. The more rapid the motion, i.e., the shorter the interval t, the more a nucleus in the molecule reacts as if it were exposed all the time to the average value of the field. It can be shown that the effect will start to become appreciable if

$$t \cong 2\pi/(\gamma \Delta H), \tag{1}$$

where γ is the gyromagnetic ratio of the nucleus under consideration. As an example we consider protons in an external field which varies over the sample by the amount $\Delta H \cong 2\times 10^{-3}$ gauss. To obtain an effective averaging of this field, one has to demand, according to Eq. (1), that the time t be less than about $\frac{1}{15}$ sec. In a sample with linear dimensions of $\frac{1}{4}$ cm this requires a speed somewhat higher than 10 cm/sec, which is certainly well within the realm of mechanical possibilities.

It must be noted that formula (1) implies that every molecule in the substance experiences the *full* range of variation ΔH during the time interval t. Generally, however, the averaging process will take place for each molecule only over those regions of the sample through which it is carried in the course of its motion. As a result one will observe a *partial* narrowing of the line which, however, can be very appreciable under suitable circumstances.

A more quantitative treatment of the effect is obtained from the phenomenological equations of nuclear induction.[2] Assuming, for example, that the motion within the substance has the effect of exposing a given molecule, after a mean time θ, to a randomly different field within the sample, one obtains for short θ a reduction of the field inhomogeneity to an "effective value"

$$(\Delta H)_{eff} = \gamma \langle (\Delta H)^2 \rangle_{Av} \theta, \tag{2}$$

where $\langle (\Delta H)^2 \rangle_{Av}$ is the mean square deviation of the field in the sample from its average value. Irrespective of the manner in which the external field varies over the sample, one obtains in this case a "Lorentzian" line with an "effective" transverse relaxation time

$$(T_2)_{eff} = 1/[1/T_2 + \gamma (\Delta H)_{eff}], \tag{3}$$

where T_2 is the transverse relaxation time due to natural broadening.

[1] This structure is distinguished from the one reported earlier by Arnold, Dharmatti, and Packard [J. Chem. Phys. 19, 507 (1951)] where the OH, CH_2, and CH_3 groups in ethyl alcohol were shown to lead to separate resonances. Higher resolution reveals that each of these resonances consists of a number of closely spaced lines. These more recent findings will soon be published.

[2] F. Bloch, Phys. Rev. 70, 460 (1946).

A Line-Narrowing Experiment

W. A. Anderson and J. T. Arnold
Stanford University, Stanford, California
(Received February 19, 1954)

THE possibility of line narrowing by macroscopic motion of the material within a sample, described in the foregoing letter, has been experimentally established. A nuclear induction apparatus consisting of a 7000-gauss permanent magnet and the associated radio-frequency components was used. The magnet was shimmed to give a homogeneity within about 10^{-3} gauss over a spherical sample of $\frac{1}{4}$-inch diameter.

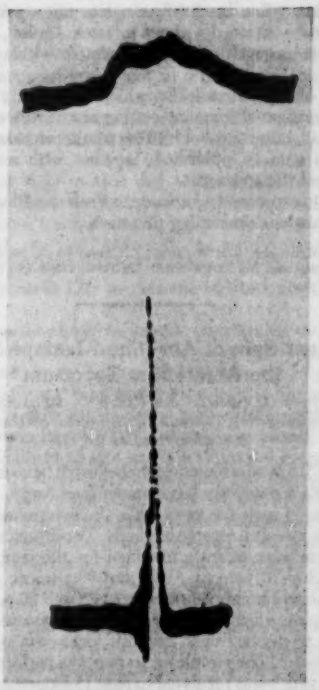

FIG. 1. Line narrowing by rotation. The top figure is a photograph of a proton resonance from distilled water displayed on the cathode-ray screen. The beam was swept from right to left in about 30 seconds. A measure of the external field inhomogeneity is given by the width at half-maximum and was found to be 1.7×10^{-3} gauss. The bottom figure was obtained by rotating the spherical sample at a rate of about 25 times per second, but under otherwise identical conditions. The variation of the field between the two ends of this trace was determined by calibration to be 3.5×10^{-3} gauss, yielding for the width at half-maximum about 10^{-4} gauss. This reduction in half-width is accompanied by an increase of the height by a factor of about 7.

A simple and effective way to induce a macroscopic motion was found to be a rotation of the entire sample about an axis (y), coincident with the axis of the receiver coil. This method achieves a partial narrowing in the sense that it effects an averaging over the field around coaxial circles in the plane normal to this axis. It has been possible in this manner to narrow the line appreciably and thereby increase its height with rotational speeds in excess of about 10 cycles per second. This is in agreement with the order of magnitude of the time t, referred to in the previous letter. The samples were enclosed in a shaft extending into the receiver coil. This shaft was driven by a small air turbine located above the head.

The figure shows oscilloscope traces of proton resonance lines in distilled water with and without rotation of the sample. It is seen that the rotation of the sample results in a reduction of the half-width by a factor of about 17 and an increase of the maximum by a factor of 7. The fact that the areas under the two curves do not seem to agree is possibly due to partial saturation in the case of the narrower line.

Since simple rotation averages only along coaxial circles, the best over-all homogeneity for the stationary samples will not necessarily represent the optimum field configuration under rotation. For example, in the best over-all field there may be a comparatively strong variation in the y direction over which no averaging takes place in the present arrangement. Thus the

relatively broad line in the figure does not represent the best resolution to be obtained without rotation. In fact, our magnet is capable of homogeneities over the sample within as little as $\frac{1}{4}$ milligauss.

With the enhanced resolution which has already been realized, we have investigated some interesting new features of the proton resonances in ethyl alcohol.[1,2] The results of these and further studies will soon be published, together with a more detailed description of the apparatus.

We wish to express our gratitude to Professor Bloch for the most helpful discussions concerning this work.

[1] Arnold, Dharmatti, and Packard, J. Chem. Phys. **19**, 507 (1951).
[2] J. T. Arnold and M. E. Packard, J. Chem. Phys. **19**, 1608 (1951).

Nuclear Spin of Am^{243} and Isotope Shift in the Americium Spectrum*

JOHN G. CONWAY AND RALPH D. MCLAUGHLIN
Radiation Laboratory, University of California, Berkeley, California
(Received February 23, 1954)

RECENT advances in nuclear chemistry have made it possible to obtain relatively large amounts of Am^{243}. The quantities available are of such size as to make the determination of nuclear spin feasible by spectroscopic methods. Two 30-microgram samples of americium were recently obtained for this purpose. One contained roughly 10 percent Am^{243} and 90 percent Am^{241}, and the other contained roughly 80 percent Am^{243} and 20 percent Am^{241}.

The samples, which were received in an aqueous solution, were evaporated to dryness upon graphite electrodes which were 0.25 inch in diameter. The sample excitation was performed by a direct current arc which was regulated to pass 15 amperes of current. The spectrum was recorded in the second and third orders of a 21-foot, 30 000 line per inch grating in a Paschen-Runge mount. An iron spectrum and an Am^{241} spectrum were photographed for comparison.

TABLE I. Isotope shifts in the americium spectrum, and wavelengths of the first component in the flag pattern of Am^{243}.

λ_{243}(A)	$\Delta\lambda_{243-241}$(A)
2815.17	+0.022
2888.46	−0.017
2909.72	+0.014
2911.06	−0.009
2938.96	+0.069
2957.13[a]	−0.009
2958.29	+0.007
2969.21	+0.014
2993.38	+0.014
3027.91	+0.021
3161.19[a]	−0.025
3204.05	+0.020
3225.27	+0.040
3257.93	−0.007
3258.62	−0.080
3282.48[a]	−0.024
3262.67[a]	−0.020
3451.99	+0.014
3473.16	+0.014
3496.48	+0.007
3926.07	+0.013
4372.35	+0.058
4441.43[a]	+0.026
4509.16	+0.032
4680.84[a]	+0.012
4699.46	+0.031
4786.64	+0.071
4871.93	+0.032

[a] Lines in which the first component is on the higher wavelength side of the pattern.

Fred and Tomkins[1] have reported the hyperfine structure of Am^{241} to be a flag pattern of six components with a spin of 5/2. No differences were observed in the hyperfine structure of Am^{243} and Am^{241}. Both structures form a flag pattern of six lines. Since Am^{243} shows six hyperfine components, it must have a nuclear spin of 5/2. There is no reason to doubt the validity of the assumption that J is greater than I. Within 2 percent error the patterns of Am^{243} and Am^{241} have the same over-all width. This indicates that the nuclear magnetic moment must be equal in the two isotopes.

The isotope shift between Am^{243} and Am^{241} was also observed. Table I contains the wavelengths of the first component of the pattern of the hyperfine lines for Am^{243} and the shift between Am^{243} and Am^{241} of that component.

We wish to thank Dr. E. K. Hulet and Dr. S. G. Thompson for providing us with the samples of americium and Professor F. A. Jenkins of the Physics Department, University of California, for the use of the spectrograph.

* This work was performed under the auspices of the U. S. Atomic Energy Commission.
[1] M. Fred and F. S. Tomkins, Phys. Rev. **89**, 318 (1953).

Angular Distribution of Photoelectrons from Gamma Radiation of 0.4 and 1.3 Mev

ARNE HEDGRAN AND SÖLVE HULTBERG
Nobel Institute of Physics, Stockholm, Sweden
(Received March 2, 1954)

THE knowledge of the angular distribution of photoelectrons and its energy dependence is of value especially for beta-spectrometric comparisons of gamma-ray intensities. At a gamma-ray energy of 2.6 Mev the angular distribution has been studied by Latyshev,[1] and the present paper gives the results of an experiment with the Au^{198} 0.41-Mev and the Co^{60} 1.33-Mev gamma rays.

The investigation was carried out by using a large double-focusing spectrometer ($\rho = 50$ cm) and with the setup shown in Fig. 1 A.

The photoelectric converter was placed at the usual source position of the spectrometer, and the gamma source was fixed to an arm that could be moved horizontally around an axis through the converter center. The setup was constructed with a minimum of material around source and converter. The photoelectron peaks were recorded with the spectrometer for different angles θ between arm and spectrometer axis. The height of the peak for an angle θ is simply proportional to the photo cross section for the corresponding unit solid angle. At larger θ the converter subtends a smaller solid angle to the source, but this effect is canceled by the fact that the gamma rays have to traverse more material in the converter. A correction has, however, to be applied because of the finite angular resolution which results from (1) geometry (the combined effect of finite spectrometer aperture, source size, and the finite solid angle under which the converter is irradiated) and (2) electron scattering in the converter.

For the Co^{60} experiment a source of 500 mC with cylindrical shape (height and diameter 4 mm) was used. The converter was a circular 7 mg/cm^2 lead foil ($\Phi = 10$ mm), source-converter distance was 45 mm, and the circular entrance aperture ($\Phi = 50$ mm) was at a distance of 300 mm from the converter. The use of two G-M counters in coincidence and elaborate shielding reduced the background from the strong source to a satisfactory level.

Figure 1 B shows the heights of photoelectron peaks with their estimated uncertainties for different θ. The angular resolution curve was calculated to be approximately a Gaussian with a full half-width of 21°, the contributions from geometry and scattering being about equal. By an unfolding procedure the experimental curve was corrected for angular resolution, giving as result the dotted curve, which should represent the true photoelectron distribution.

For the Au^{198} experiment a sufficiently thin lead converter was not available, so a gold foil (1.7 mg/cm^2) was used as converter. A 400-mC source (diameter 5 mm, height 1 mm) enclosed in aluminum was used, and the geometry was similar to that in the previous experiment. Figure 1 C shows the observed photoline peak heights plotted against θ (full line) and the corrected curve

(dotted line). The angular resolution curve had in this case a full half-width of about 20°.

Sauter[2] has derived an expression for the angular distribution of photoelectrons at relativistic energies, but his result is questionable for photoeffect in heavy elements. The agreement between the present results and his formula seems, however, to be fairly good.

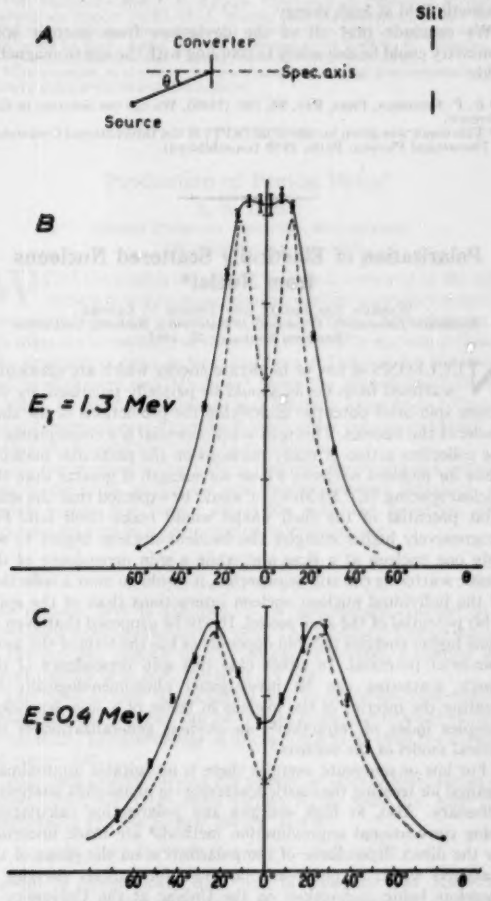

FIG. 1. A. Experimental setup (schematic). B, C. Angular distributions of photoelectrons from gamma radiations of 1.3 and 0.41 Mev, respectively.

According to him the maxima in the distributions for 0.4 and 1.3 Mev should occur for $\theta=28°$ and 12°, and we find them for $\theta=24°$ and 12°, respectively.

[1] G. D. Latyshev, Revs. Modern Phys. 19, 132 (1947).
[2] F. Sauter, Ann. Physik 9, 217 (1931).

Production of Antiprotons in p-p Collisions

DAVID FOX*
Department of Physics, University of California, Berkeley, California
(Received January 15, 1954)

THE cross section for the production of a proton pair by the collision of two protons has been calculated in the neighborhood of threshold, using the formalism suggested by Feynman for the solution of such problems.[1]

The Feynman diagrams for the process are topologically similar to those for the triple scattering of protons. In the latter case, there are 36 lowest-(fourth) order diagrams, each of which consists of three proton paths, with a meson exchange between the first and second protons, and, at a later proper time along the second path, another meson exchange between the second and third protons. The 36 diagrams are obtained from any standard one by the permutations of the three initial particles and of the three final ones. The process under consideration differs from triple scattering only in that one of the initial particles is the emitted antiproton; in each diagram the path segment representing the antiproton has a reversed proper-time sense.

At threshold, the total energy in the center-of-mass frame is $4M$, where M is the rest mass of the proton. In the laboratory system, this becomes $8M$, so that the incident kinetic energy is $6M=5.6$ Bev.

Some information about the angular distribution and the energy dependence near threshold may be obtained from symmetry arguments alone. From the fact that at least two of the emitted positive protons must have the same spin, one obtains a simple condition on the momentum dependence of the matrix element. It may be shown in this way that if the matrix element corresponding to any assignment of spin states is expanded in powers of the 3-moments of the four emitted particles, the leading terms are linear in the momentum components of the protons and are independent of the momentum of the antiproton. Therefore, (except for the possible occurrence of certain relationships among the numerical coefficients) the antiproton will be emitted spherically symmetrically in the center-of-mass system, while the protons will not be.

The energy dependence of the total cross section may be determined, except for a constant factor, by examining the form of the integral over momentum space. The available volume in momentum space is proportional to $(E-2M)^{7/2}$, where E is the total energy of either incident proton in the center-of-mass frame.[3] It is easily shown that the leading terms of the differential cross section contribute a factor proportional to $(E-2M)$ after integration; that is, the total cross section near threshold is of the form

$$\sigma=K\left(\frac{g^2}{\pi}\right)^4\left(\frac{E-2M}{M}\right)^{5/2}.$$

where the coupling constant is defined as in reference 1, with $\hbar=c=1$. The arguments used to obtain this equation are valid for all the usual weak-coupling theories, so that, in general, one can expect that only the value of K will differ from one theory to another in this approximation. K was calculated only for the case of pseudoscalar mesons with pseudovector coupling; the result was $K=1.2\times10^{-35}$ cm^2.

This cross section may be compared with that obtained by Fermi[3] in his statistical theory of high-energy nuclear processes:

$$\sigma=7\times10^{-28}\left(\frac{E-2M}{M}\right)^{7/2}.$$

In this case, the energy dependence arises from the volume of momentum space alone, since the matrix element is assumed to be a constant. Fermi's theory is a strong-coupling theory, so that to a certain extent the two results represent opposite limiting cases.

Taketani and Machida[4] have calculated the cross section for the same process, using the Feynman method and the PS-PS theory, but they obtained a fourth power dependence on $(E-2M)$. However, the details of their calculations are not clear, and there is some question as to whether certain of their approximations are justified.

Further details will be published elsewhere.

The author wishes to acknowledge his indebtedness to Professor Robert Serber for suggesting the problem and for his helpful and stimulating discussions.

* Now at the Israel Institute of Technology, Haifa, Israel.
[1] R. P. Feynman, Phys. Rev. 76, 749, 769 (1949).
[3] E. Fermi, Progr. Theoret. Phys. (Japan) 5, 570 (1950).
[4] M. Taketani and S. Machida, Progr. Theoret. Phys. (Japan) 6, 339 (1951).

Proton-Neutron Mass Difference

R. P. FEYNMAN AND G. SPEISMAN
California Institute of Technology, Pasadena, California
(Received February 23, 1954)

SUPPOSE all deviations from isotopic spin symmetry are due solely to electromagnetic effects. Then such things as the mass difference of charged and neutral π mesons, and the neutron-proton mass difference would have to be just electrodynamic.

We have investigated this point and have found that it is a reasonable possibility. For particles of zero spin, like π mesons, assumed to be elementary particles, the self-energy is quadratically divergent. If the photon propagation function[1] $1/k^2$ is cut off by a convergence factor $C(k^2)=[\Lambda^2/(\Lambda^2-k^2)]^2$, the resulting energy is about $3e^2\Lambda^2/8\pi m$, where Λ is the cut-off energy and m is the π-meson mass, and we assumed $\Lambda \gg m$. This gives the observed mass difference of about 11 electron masses for a cut-off Λ of about 1.0 proton masses.[2]

It is usually assumed that the negative value of the proton-neutron mass difference speaks against an ultimate electromagnetic explanation. The following calculation shows that this is an unwarranted assumption. The particles are not simple and there is uncertainty as to the correct law of coupling to the electromagnetic field. But for low energy, the proton can be represented by the Dirac equation with an additional Pauli term to represent the anomalous moment. Since we do not know to how high an energy this may be a reasonable approximation, we have tried providing the moment coupling term with a cut-off factor of its own. We write for the proton self-energy

$$\Delta M=(e^2/\pi i)\int\left[\gamma_\mu-\frac{\mu}{4M}(\gamma_\mu k-k\gamma_\mu)G(k)\right](\not p-\not k-M)^{-1}\times\left[\gamma_\mu+\frac{\mu}{4M}(\gamma_\mu k-k\gamma_\mu)G(k)\right]k^{-2}d^4kC(k)$$

in the notation of reference 1. We used $G(k)=-\lambda^2(k^2-\lambda^2)^{-1}$ to cut the moment coupling off at energies about λ, and $C(k)=-\Lambda^2(k^2-\Lambda^2)^{-1}$ to cut off the photon propagation function at energy Λ. The expression for the neutron is the same, except that the γ_μ coupling terms are omitted and the value of μ, the anomalous moment in nuclear magnetons, is -1.91 instead of 1.79 for the proton. M is the nucleon mass.

For the proton the term for $\mu=0$, representing coupling of current with current, is positive, as is also the term in μ^2. But the cross term, linear in μ, is negative and quite large if the moment is not cut off too soon. Thus the proton-neutron mass difference can easily turn out negative. For example, if Λ and λ are both taken at $1.4M$, the experimental value of -2.5 electron masses results for this difference (in this case for the neutron ΔM is roughly 1.5, for the proton -1.0 electron masses). No small difference of large numbers is involved. If the cut-off λ is reduced below about $0.75M$, a negative mass difference cannot be obtained. For $\lambda=1.0M$, $\Lambda=4M$ gives the experimental result.

The high cutoff for the anomalous moment implies that the charge responsible for the moment must be spread over only a small distance (of order $\hbar/Mc$). This is also suggested by the relatively small changes that the nucleon moments undergo when nucleons form nuclei.

The cutoff for the propagation function may be interpreted in two ways. Firstly, electrodynamics may fail at high energies, the failure being represented in a crude way by the cutoff. If this is so we could guess from our results that the failure occurs at energies in the neighborhood of the nucleon mass. Another possibility is that the electrodynamics is correct, but the cutoff represents, roughly, the error committed in assuming that the particles are elementary. For example, in the case of the π meson, we have assumed the π meson in virtual states acts as a simple particle. But for energies as high as M, strongly coupled virtual nucleon pairs may be formed. They, rather than failure of electrodynamics, may provide the convergence at energies of order M. In a like manner, the complex of virtual mesons presumed to be associated with nucleons may have the effect that at sufficiently high energy the electromagnetic coupling of neutron and proton may be nearly the same, so that the integral representing the difference of their masses may converge without modification of electrodynamics. In this way, the presumed convergence of the mass differences might tell us something about the character of coupling with the electromagnetic field at high energy.

We conclude that all of the deviations from isotopic spin symmetry could be due solely to coupling with the electromagnetic field.

[1] R. P. Feynman, Phys. Rev. 76, 769 (1949). We use the notation in this reference.
[2] This result was given by one of us (RPF) at the International Conference in Theoretical Physics, Paris, 1950 (unpublished).

Polarization of Elastically Scattered Nucleons from Nuclei*

WARREN HECKROTTE AND JOSEPH V. LEPORE
Radiation Laboratory, University of California, Berkeley, California
(Received February 23, 1954)

NUCLEONS of low or moderate energy which are elastically scattered from nuclei should be partially polarized[1] by the strong spin-orbit potential underlying the predictions of the shell model of the nucleus. This spin-orbit potential is a consequence of the collective action of many nucleons on the particular nucleon. Thus for incident nucleons whose wavelength is greater than the nuclear spacing ($E \lesssim 50$ Mev), it would be expected that the spin-orbit potential of the shell model would make itself felt. For progressively higher energies the incident nucleon begins to see only one nucleon at a time and while a spin dependence of the elastic scattering can still be expected, it would be more a reflection of the individual nucleon-nucleon interactions than of the spin-orbit potential of the shell model. It will be supposed that even at these higher energies the spin dependence has the form of the usual spin-orbit potential. In either case this spin dependence of the elastic scattering can be investigated phenomenologically by treating the interior of the nucleus in terms of a spin-dependent complex index of refraction[2]—an obvious generalization of the optical model of the nucleus.[3]

For low or moderate energies there is no suitable approximate method for treating the elastic scattering—a phase-shift analysis is necessary. Also, at high energies any polarization calculations using conventional approximation methods[4] are made uncertain by the direct dependence of the polarization on the phase of the scattered wave. A phase-shift analysis for various energies is therefore being undertaken on the Univac at the University of California Radiation Laboratory at Livermore in collaboration with S. Fernbach.

An estimate for small angles of scattering, though rough at best, may be readily obtained by making several simplifying assumptions. The magnitude of the polarization is given by

$$P=\left(\frac{AB^*+A^*B}{d\sigma/d\Omega}\right)\sin\theta. \qquad (1)$$

Here A and B represent the scattering amplitudes corresponding to the spin-independent and spin-dependent parts of the interaction, respectively. The known experimental value for the differential cross section $d\sigma/d\Omega$ may be used. The amplitude, B, for spin-dependent scattering may be estimated by using the Born approximation. Then only the imaginary part of A contributes to P. For small angles this is approximately proportional to the total cross section.

For 300-Mev neutrons incident on carbon, for example, a square-well spin-orbit interaction ($R=1.4A^{\frac{1}{3}}\times10^{-13}$ cm) of 2-Mev depth gives a polarization of 40 percent at five degrees. Though this is probably an overestimate, it suggests that the existence of a small

spin-orbit interaction in the nuclear potential will provide large polarizations. Finally, aside from the interest of investigating one more phase of nuclear structure, this phenomenon, because of the large diffraction cross sections, provides a relatively good source of polarized particles.

* This work was performed under the auspices of the U. S. Atomic Energy Commission.

[1] W. Heckrotte and J. V. Lepore, University of California Radiation Laboratory Report UCRL-2152, March, 1953 (unpublished), p. 32.

[2] This method has been suggested in connection with low-energy neutron scattering by R. K. Adair (private communication from K. M. Watson).

[3] Fernbach, Serber, and Taylor, Phys. Rev. 75, 1352 (1949).

[4] For example, in the special case of real potentials the Born approximation always gives a vanishing polarization.

Production of Proton Pairs*

R. N. Thorn†
Harvard University, Cambridge, Massachusetts
(Received February 1, 1954)

WITH the completion of the California bevatron in the near future it will be energetically possible to produce negative protons, if indeed they exist; and it is therefore of some interest to calculate the probability of such an occurence by nucleon collision. The production cross section has been calculated by Helstrom[1] and by Taketani and Machida,[2] using the Weizsäcker-Williams method. In addition, Taketani and Machida have made a threshold approximation of Feynman's method for the production of an antiproton by proton-proton $(p,\bar{p})$ collision.

It is the purpose of this note to describe the results of a calculation of the cross section for pair production near threshold, an energy region just barely attainable with the bevatron. The processes treated in a perturbation calculation are:

$$p+p\rightarrow p+p+p+\bar{p},$$
$$p+n\rightarrow p+n+p+\bar{p}.$$

The nucleon field is assumed to be pseudoscalar-coupled to a symmetric pseudoscalar meson field. The threshold kinetic energy of the incident particle in the laboratory system, assuming the target particle is at rest, is $6M$ or 5.6 Bev.

The cross section is calculated in the center-of-mass system where the final particles can be treated nonrelativistically so long as the initial particle energy is not much greater than the threshold energy $2M$.

Since only neutral mesons are present in intermediate states to lowest order, the $(p,\bar{p})$ cross section is the same for both the symmetrical and the neutral theory, and is[3]

$$\phi(p,\bar{p})=1.4\times10^{-29}(f^2/4\pi)^4[(E-2M)/M]^{7/2}\ \text{cm}^2,$$

a result in disagreement with that of Taketani and Machida. Here $E-2M$ is the excess over threshold energy of an initial particle in the center-of-mass system.

The $(p,\bar{p})$ production process is inhibited by the exclusion principle, since it is unallowable to have three final protons in a state of zero momentum. This restriction no longer applies if one of the final particles is a neutron, as in the case for antiproton production by $(n,\bar{p})$ collision. Furthermore, in a symmetric theory there are a greater number of processes contributing because of the possibility of charged as well as neutral mesons in intermediate states. These two factors together make the $(n,\bar{p})$ production predominant over the $(p,\bar{p})$ production near threshold.

The $(n,\bar{p})$ cross section is

$$\phi(n,\bar{p})=5.4\times10^{-29}(f^2/4\pi)^4[(E-2M)/M]^{7/2}\ \text{cm}^2.$$

I wish to thank Professor Schwinger for having suggested this problem.

* Based on a thesis written in partial fulfillment of the requirements for a doctor's degree.

† Now at Los Alamos Scientific Laboratory, Los Alamos, New Mexico.

[1] C. W. Helstrom, Phys. Rev. 78, 88 (1950).

[2] M. Taketani and S. Machida, Prog. Theoret. Phys. 6, 539 (1951).

[3] Units $\hbar=c=1$ have been used, so that $f^2/4\pi$ is the usual dimensionless fine structure constant of meson theory.

Summation of γ-Ray Energies with a Single-Crystal Spectrometer

Daniel C. Lu and M. L. Wiedenbeck
Randall Laboratory of Physics, University of Michigan, Ann Arbor, Michigan
(Received February 19, 1954)

MULTI-γ coincidences were effectively observed with a simple modification of a scintillation spectrometer. This was accomplished by drilling a hole 3 mm in diameter along the axis of a cylindrical NaI(Tl) crystal from the top surface to the center. The bottom surface was coupled to a photomultiplier in the usual manner. This allowed the source to be placed outside of the crystal or at its center. With the source within the crystal, there is a reasonable probability that all prompt, cascade γ rays from a single disintegration will be absorbed and thus give a pulse corresponding to the sum of the γ energies involved. As such this device can often yield data not readily obtained with the conventional arrangement of spectrometers in coincidence.

The scheme was first tried some months ago using a crystal 1 inch thick and 1¾ inches in diameter. While this showed that the system could yield useful information, it was inadequate for higher energy γ rays. At present we use a crystal 2 inches in diameter and 2 inches thick coupled to a DuMont 6292 phototube. Pulses are observed with a single-channel discriminator.

Figure 1 shows the spectrum of γ rays following the β decay of Co^{60} taken with the source outside the crystal (with small probability for the addition of γ-ray energies) as well as the spectrum taken with the source at the center of the 2-inch crystal. Even at these relatively high energies (1.17 and 1.33 Mev) the sum peak is easily observed with this crystal.

The γ-ray spectra following the β decay of Cs^{134} is indicated in Fig. 2. The main sum peaks are due to triple "coincidences" be-

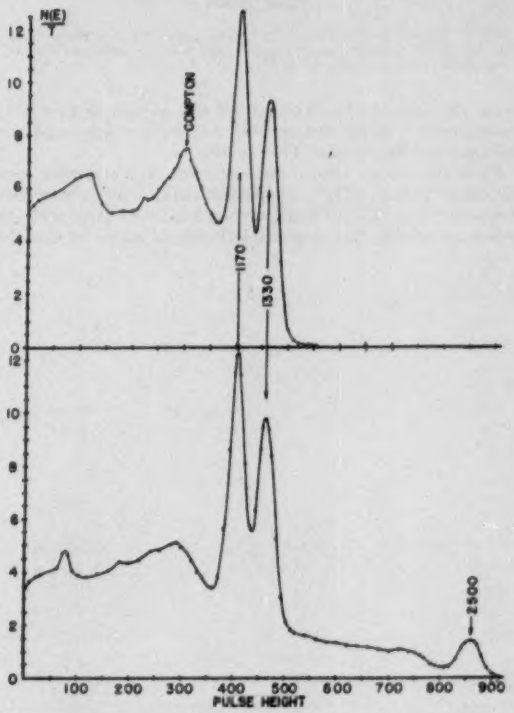

FIG. 1. Gamma-ray spectra of Ni^{60}. Energies in kev indicated at the top of peaks. Upper curve with sample 3 in. from top of crystal. Lower curve with sample inside crystal. The peak representing the sum of the two gamma energies is seen at 2.5 Mev.

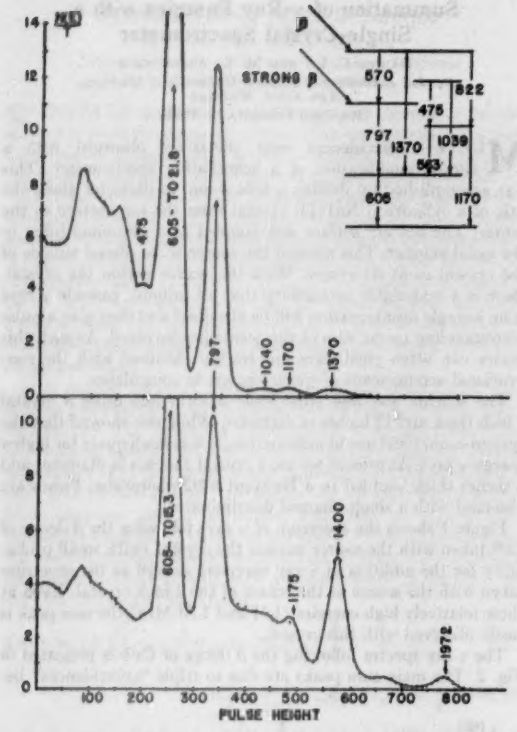

FIG. 2. Gamma-ray spectra of Ba^{134}. The upper curve is the ordinary spectrum. The lower curve shows several peaks due to the addition of two or three gamma rays. Energies are in kev.

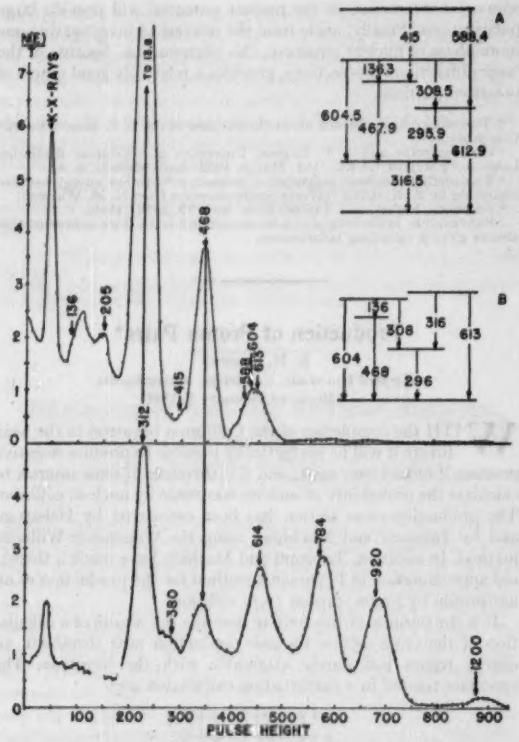

FIG. 3. Gamma spectra of Pt^{193} with proposed decay schemes. The present data favor the scheme *A*. Energies are in kev.

tween the lines of 570, 797, and 605 kev as well as to double "coincidences" of the strong 797- and 605-kev lines and the 605-kev line with the weak 570-kev line.

From the precise energy measurements of the gamma rays following the decay of Ir^{193}, two possible decay schemes have been proposed.[1,2] (See Fig. 3.) The spectrum was investigated with the source outside the crystal as well as inside. It is evident that the highest sum peak corresponds to 1.2 Mev and peaks are also found at 920 and 794 kev. In addition the peak at 614 kev is greatly augmented. It is clear that these data are in agreement with decay scheme *A* but not with *B*.

[1] Cork, LeBlanc, Stoddard, Childs, Branyan, and Martin, Phys. Rev. **82**, 258 (1951).
[2] Muller, Hoyt, Klein, and DuMond, Phys. Rev. **88**, 775 (1952).